NTP
National Toxicology Program

Review of the Biomolecular Screening Branch

by the

NTP Board of Scientific Counselors

November 30 – December 1, 2010

Rodbell Auditorium, Rall Building
National Institute of Environmental Health Sciences
Research Triangle Park, NC

This page left intentionally blank

Table of Contents

Review of the Biomolecular Screening Branch by the NTP Board of Scientific Counselors

List of Abbreviations and Acronyms

Affiliations/Organizations

ATSDR	Agency for Toxic Substances and Disease Registry
BB	Biostatistics Branch
BSB	Biomolecular Screening Branch
CEINT	Center for the Environmental Implications of NanoTechnology
CERHR	Center for the Evaluation of Risks to Human Reproduction
CMPB	Comparative Medicine and Pathology Branch
CVM	Center for Veterinary Medicine
DERT	Division of Extramural Research and Training
DIR	Division of Intramural Research
IATAP	Interagency Alternative Technology Assessment Program
IHCP	Institute for Health and Consumer Protection
IRG	Intracellular Regulation Group
LMC	Laboratory of Molecular Carcinogenesis
LN	Laboratory of Neurobiology
LTP	Laboratory of Toxicology & Pharmacology
LRB	Laboratory of Respiratory Biology
LRDT	Laboratory of Reproductive & Developmental Toxicology
LSB	Laboratory of Structural Biology
LST	Laboratory of Signal Transduction
MLPCN	Molecular Libraries Probe Production Centers Network
MCG	Medicinal Chemistry Group
NAS	National Academy of Sciences
NCCT	National Center for Computational Toxicology
NCEH	National Center for Environmental Health
NCGC	NIH Chemical Genomics Center
NCTR	National Center for Toxicological Research
NHEERL	National Health and Environmental Effects Research Laboratory
NHGRI	National Human Genome Research Institute
NICEATM	NTP Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NIST	National Institute for Standards and Technology
NLB	NTP Laboratories Branch
NRC	National Research Council
NTP	National Toxicology Program
POB	Program Operations Branch
ORD	U.S. EPA Office of Research and Development
STAR	U.S. EPA CompTox Science to Achieve Results
TB	Toxicology Branch
TRND	Therapeutics for Rare and Neglected Diseases
UNC-CH	University of North Carolina at Chapel Hill
U.S. DOD	U.S. Department of Defense

U.S. EPA	U.S. Environmental Protection Agency
U.S. FDA	U.S. Food and Drug Administration

Gene Names

Abcb1b	ATP-binding cassette subfamily B member 1b
Abcc3	ATP-binding cassette subfamily C member c
Adam8	A Disintegrin and Metalloprotease domain family member 8
AhR	aryl hydrocarbon receptor
Akt	v-akt murine thymoma viral oncogene homolog 1
AR	androgen receptor
c8orf46	chromosome 8 open reading frame 46 homolog
CAR	constitutive androstane receptor
CAsE-PE	chronic-arsenic–exposed human prostate epithelial cells
CCL2	chemokine (C-C motif) ligand 2
Cdh13	cadherin 13, H-cadherin
Cinc1	cytokine induced neutrophil chemoattractant 1
CREB	cAMP response element binding transcription factor
Ddit4L	DNA-damage-inducible transcript 4-like
ERα	estrogen receptor alpha
ERβ	estrogen receptor beta
ERK	mitogen-activated protein kinase 1
Fhit	fragile histidine triad gene
FXR	farnesoid X receptor
GR	glucocorticoid receptor
Grin2c	glutamate receptor ionotropic, N-methyl D-aspartate 2C
h	human (in the context of gene names)
HMG-CoA	3-hydroxy-3-methylglutaryl-Coenzyme A
HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase (mitochondrial)
LXR	human liver X receptor
MAD1	mitotic arrest deficient 1
MAD2	mitotic arrest deficient 2
MBD2	Methyl-CpG-binding domain protein 2
Mybl2	v-myb meloblastosis viral oncogene homolog (avian)-like 2
PPARα	human peroxisome proliferator activator receptor alpha
PPARγ	human peroxisome proliferator activator receptor gamma
PPARδ	human peroxisome proliferator activator receptor delta
PR	human progesterone receptor
PXR	pregnane X receptor
RAR	human retinoic acid receptor
RXR	human retinoid X receptor
TRα	thyroid hormone receptor alpha
TRβ	human thyroid hormone receptor beta
VDR	human vitamin D receptor
RASSF1A	RAS association family 1 gene
TFF3	trefoil factor 3
Wwox	WW domain-containing oxidoreductase

Abbreviations

3'Seq	3 prime based-next generation sequencing of RNA
A	Angstrom
AA	androgen antagonism
Abs	absorbance
ACToR	Aggregated Computational Toxicology Resource)
AFB_1	Aflatoxin B1
ATP	adenosine triphosphate
BPA	Bisphenol A
CAS	Chemical Abstracts Service
CEBS	Chemical Effects in Biological Systems
CGH	comparative genomic hybridization
CLND	chemiluminescent nitrogen detection
Comp Tox	U.S. EPA Computational Toxicology Research Program
CNV	copy number variation
COG	Clusters of Orthogonal Groups of proteins (database)
CSV	comma-separated values
Ct	cycle threshold
CTPE	cadmium-transformed prostate epithelial cells
CTD	Comparative Toxicogenomics Database
dDEV	fetal effects (weight reduction, defects)
DMSO	dimethyl sulfoxide
DSSTox	Distributed Structure-Searchable Toxicity Database
EC_{50}	effective concentration of a chemical that give half-maximal response
ELISA	enzyme-linked immunosorbent assay
ELSD	evaporative light scattering detection
ExpoCast	Exposure Forecaster
EXT	extinction
FFPE	formalin fixed, paraffin embedded
Fl	fluorescent intensity
FP	fluorescent polarization
FRET	Förster resonance energy transfer
GFP	green fluorescent protein
GO	Gene Ontology
h	hour (in the context of time)
HCC	hepatocellular carcinomas
HPV	human papilloma virus
HTS	high throughput screening
IC_{50}	Inhibition concentration that induces a half-maximal decrease in response
kB	kilobases
KEGG	Kyoto Encyclopedia of Genes and Genomes
kNN	k nearest neighbor
LBD	ligand binding domain
LEC	lowest effective concentration
LEL	lowest effect level
LOPAC	Library of Pharmacologically Active Compounds
mDEV	maternal effect (weight gain, pregnancy loss)

mM	millimolar
MMTV	mouse mammary tumor virus
MW	molecular weight
nM	nanomolar
nm	nanometers (do you want this spelled out in the text?)
Pgp	p-glycoprotein
qNPA	quantitative nuclease protection assay
OS	oxidative stress
Pa	Pascal (a unit of pressure equal to one newton per square meter)
q	quantitative
QC	quality control
QSAR	quantitative structure-activity relationship
RBF	radial basis function
RNAi	RNA interference
RNA-Seq	Next generation sequencing of RNA
RT-PCR	real-time polymerase chain reaction
RWPE-1	immortalized human prostatic epithelial cell line
SBIR	Small Business Innovative Research
SEE	Senior Environmental Enrollees
siRNA	small inhibitory RNA
SNP	single nucleotide polymorphism
SVM	support vector machine
TOF	time of flight
TMT	trimethyltin
ToxCast™	Toxicity Forecaster
ToxMiner™	Interface for visualizing and analyzing ToxCast™ data
ToxPi	Toxicological Priority Index
ToxRefDB	Toxicity Reference Database
TRF	time resolved fluoresence
TSV	tab-separated values
µM	micromolar
WFS	weighted feature significance
XML	extensible markup language

I. Overview of the BSB and the Tox21 Initiative

This page left intentionally blank

I. Overview of the BSB and the Tox21 Initiative

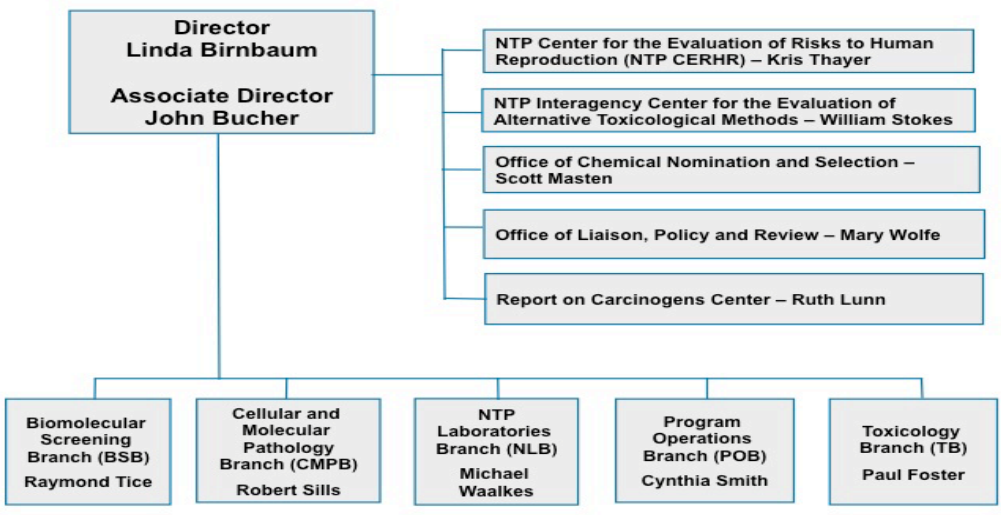
Raymond R. Tice, Ph.D.
Chief, Biomolecular Screening Branch

I.1 Background

I.1.1 Initial Efforts (2005 to 2007)

The NTP Vision for the 21st Century (http://ntp.niehs.nih.gov/go/vision, "the Vision") is to move toxicology from a predominantly observational science at the level of disease-specific models to a predominantly predictive science focused upon a broad inclusion of target-specific, mechanism-based, biological observations. To implement the Vision, the NTP developed a Roadmap (http://ntp.niehs.nih.gov/go/vision, "the Roadmap"), made public in March 2005, that places an increased emphasis on the use of alternative assays for targeting the key pathways, molecular events, or processes linked to disease or injury, and attempts to incorporate them into a research and testing framework. The Roadmap positions the NTP to provide scientific data and the interpretation of those data for public health decision-making. As a logical outgrowth of the Roadmap, NTP established a High Throughput Screening (HTS) program, representing a new paradigm in toxicological testing. The HTS program's approach is to develop and/or identify toxicological testing screens for mechanistic targets active within cellular pathways considered critical to adverse health effects (i.e., so called "toxicity pathways"), such as carcinogenicity, reproductive and developmental toxicity, genotoxicity, neurotoxicity, and immunotoxicity, in humans. The Biomolecular Screening Branch (BSB; http://www.niehs.nih.gov/research/atniehs/labs/bmsb/index.cfm), established in 2007 within the NIEHS/NTP, administers the HTS program (**Figure I.-1**).

Figure I.-1 Pending organizational structure of the National Toxicology Program within the National Institute of Environmental Health Sciences



Beginning in 2005, the NTP made a number of critically important steps to ensure success of the HTS program. In May 2005, the NTP entered into collaboration with the recently established NIH Chemical Genomics Center, a member of the NIH Molecular Libraries Screening Center Network (MLSCN) (http://mli.nih.gov/mli/), to screen environmental compounds for biological activity in biochemical- and cell-based quantitative (q)HTS assays[1] in 1536-well plates using robotics. The MLSCN was initiated on the heels of the Human Genome Project to produce chemical probes of biological systems to deconstruct the complexity of the genome and study the principles by which chemical and biological space interact. However, the NTP and the NCGC both recognized that the same approach might be used for the *in vitro* toxicological investigation of the thousands of compounds to which humans are exposed.

By December 2005, the NTP had provided the NCGC a library of 1408[2] compounds (1353 unique, 55 duplicate) as well as the first set of *in vitro* assays to be used for qHTS. Compound selection was based primarily on having been tested by the NTP in one or more toxicological assays. The *in vitro* assays (obtained from Promega) included two for measuring cytotoxicity, three for evaluating the induction of different caspases (as a measure of the ability of compounds to induce apoptosis), and one for measuring effects on the p-glycoprotein (Pgp) transporter involved in drug resistance. After optimizing the assays for a 1536-well format, qHTS began in February 2006. In 2007, in recognition of the importance of this collaboration to the HTS program, the NTP entered into a five-year Interagency Agreement "Toxicity Profiling Using High Throughput Screening" with the NCGC to directly support the expanding effort.

Also, in December 2005, the NTP sponsored the High Throughput Screening Assays Workshop (http://ntp.niehs.nih.gov/go/13038), held in Arlington, VA, the objectives of which were to:

- educate the NTP on how HTS assays are typically conducted
- identify which HTS assays might be the most informative in terms of possibly predicting toxicological responses in laboratory animals and humans
- recommend how compounds might be selected/prioritized for testing in HTS assays
- describe how the HTS assay data should be collected, stored, and mined
- consider how U.S. Federal agencies could use information from these assays in making regulatory decisions.

The recommendations resulting from this workshop have guided the decisions made by NTP relating to its HTS program.

By 2006, it became apparent that the NTP's collaboration with the NCGC and information on its qHTS capabilities shared some common goals and approaches with the recently established U.S. EPA National Center for Computational Toxicology (NCCT) (http://www.epa.gov/ncct/). As a result, later that year, the NCCT arranged also for the screening of a library of 1408 compounds of interest to the U.S. EPA at the NCGC and identified additional *in vitro* assays of interest that would be used for screening the NTP

[1] qHTS is a titration-based screening approach that efficiently identifies biological activities in large compound libraries

[2] The reason for 1408 compounds in that in the NCGC qHTS strategy, a 1536-well plate has 1408 wells available for compounds, the remaining 128 wells in the first four columns are used for negative and positive controls; when screening multiple concentrations with different plates representing different concentrations.

library as well. The U.S. EPA library was later expanded by 54 compounds to meet the needs of the NCCT related to compounds being screened for biological activity in the ToxCastTM Program (http://www.epa.gov/ncct/toxcast/). ToxCastTM, launched by the U.S. EPA in 2007, involved the screening of 320 predominantly pesticide active compounds (309 unique) in over 500 in vitro assays via a number of contracts and agreements. Similar to the HTS Initiative for the NTP, the purpose of ToxCastTM is to revolutionize the U.S. EPA's chemical toxicity evaluation procedures by using advances in computers, genomics, and cellular biology to accelerate toxicity testing and enhance the agency's capacity to screen new compounds. Early on, the NTP and the U.S. EPA recognized that their efforts were complementary and decided that the NTP would focus on the qHTS screening of extensive compound libraries at the NCGC while the EPA would focus on screening smaller numbers of compounds across mid- and high-throughput assays covering greater biological diversity.

I.2 The National Academy of Sciences Report "Toxicity Testing in the 21st Century: A Vision and Strategy"

Independent of these activities, in 2007, the National Academy of Sciences (NAS) published the National Research Council (NRC) report "Toxicity Testing in the 21st Century: A Vision and Strategy" (http://www.nap.edu/catalog.php?record_id=11970) that envisioned a not-so-distant future in which virtually all routine toxicity testing would be conducted *in vitro* in human cells or cell lines by evaluating perturbations of cellular responses in a suite of toxicity pathway assays using high throughput robotic-assisted methodologies. This report raised concerns about the current largely animal-based approach for toxicity characterization and expressed the desire/need to reduce the number of animals used in testing, to reduce the overall cost and time required to characterize each compound, and to increase the level of mechanistic understanding of compound toxicity. This NRC committee effort was supported by the U.S. EPA, with additional funding from the NTP.

The phased approach for toxicity determination outlined by the NRC report includes compound characterization (e.g., physico-chemical properties, metabolism, environmental distribution, exposure risk), followed by "toxicity pathway" characterization using biochemical and cell-based *in vitro* tests to indicate which (if any) "toxicity pathways" are activated by the test compound. For a subset of compounds, "targeted testing" (e.g., additional *in vitro* assays or focused *in vivo* animal testing) would be carried out to better characterize the risk for human toxicity. The final phase is "dose response and extrapolation modeling" to perform low dose extrapolation, toxicokinetics, and exposure estimation. All phases would have significant computer modeling components. The outcome of these phases would be a determination of the potential toxic effects (including mode and mechanism of action) of a compound, as well as estimates of dose response behavior.

The premise behind the NRC vision is that the effect of a compound is ultimately due to direct or indirect molecular interactions with one or more cellular component(s) (e.g., receptors, mitochondria, DNA) that excessively perturbs a toxicity pathway, leading to an adverse health outcome. However, for many reasons (e.g., the need to identify all toxicity pathways; the availability of appropriate *in vitro* assays for each toxicity pathway; the ability to incorporate human relevant xenobiotic metabolism into *in vitro* assays, to extrapolate *in vitro* concentration to *in vivo* dose, and accurately predict absorption, distribution, excretion, and metabolism in humans), it is unlikely that this vision can be fully achieved without the commitment of very extensive scientific and financial resources and significant scientific advances. Despite this, it should be possible to implement, in a much shorter time and with less cost, a generalized set of assays for key biological targets and pathways that can be used for the purpose of initial compound screening and prioritization.

I.3 Tox21: A Federal Partnership Transforming Toxicology

In recognition that a coordinated program among multiple government organizations would be needed to make progress on the NRC vision (by building on existing expertise and overcoming the resource limitations of a single agency), the NTP, NCGC, and the U.S. EPA entered into a formal partnership to screen a large number of compounds broadly characterizing and defining the chemical-biological space occupied by compounds of toxicological concern. This partnership was announced in a five-year memorandum of understanding (MOU) released on February 14, 2008, along with a Science publication authored by Drs. Francis Collins (NHGRI), George Gray (EPA), and John Bucher (NIEHS/NTP) describing how this partnership would transform environmental health protection (Collins et al. 2008). The MOU builds on the experimental toxicology expertise at the NTP, the qHTS technology of the NCGC, and the computational toxicology capabilities of the NCCT. This year, on July 19, a new five-year MOU (http://ntp.niehs.nih.gov/go/28213) was announced that expanded the collaboration to include the U.S. Food and Drug Administration (FDA). The U.S. FDA brings to the partnership its experience in human diseases and in animal models of human disease, as well as in toxicity pathway analysis and computational toxicology. Thus, each Tox21 partner brings complementary expertise to bear on the application of novel methodologies to evaluate large numbers of compounds for their potential to interact with biological processes relevant to toxicity, as well as on the analysis and interpretation of the resulting data.

In the Tox21 MOU, the agencies agree to work together to:

- research, develop, validate, and translate innovative compound testing methods that characterize toxicity pathways
- identify compounds, assays, informatic analyses, and targeted testing needed for the innovative testing methods
- prioritize compounds for more extensive toxicological evaluation
- identify mechanisms of compound-induced biological activity in order to characterize toxicity pathways, facilitate cross-species extrapolation, and provide input to models for low-dose extrapolation
- develop predictive models for biological response in humans

Consistent with the vision outlined in the NRC report, success in achieving these goals is expected to result in methods for toxicity testing that are more scientifically based and cost effective as well as models for risk assessment that are more mechanistically based. As a consequence, a reduction or replacement of animals in regulatory testing is anticipated to occur in parallel with an increased ability to evaluate the large numbers of compounds that currently lack adequate toxicological evaluation. Ultimately, Tox21 is expected to deliver biological activity profiles that are predictive of *in vivo* toxicities for the thousands of under-evaluated substances of concern to regulatory authorities not only in the United States, but in other countries as well.

To support the goals of Tox21, four working groups—Compound Selection, Assays and Pathways, Informatics, and Targeted Testing—have been established; these working groups represent the different components of the NRC vision. The purpose and activities of each group are described in Section III of this background document. In addition to the testing activities, the MOU promotes coordination and sponsorship of workshops, symposia, and seminars to educate the various stakeholder groups, including

regulatory scientists and the public, with regard to Tox21-related activities. Since the original MOU was made public, representatives of the NTP, NCGC, and the U.S. EPA have made almost 200 invited presentations (e.g., seminars, conference platform and posters presentations) connected with Tox21 at national and international conferences. As one important example, I gave the opening plenary presentation "The U.S. Tox21 Community and the Future of Toxicology Testing" at the VII World Congress on Alternatives and Animal Use in the Life Sciences: Calling on Science, held in Rome, Italy in late August, 2009.

To coordinate Tox21 activities, there is a primary point of contact for each agency (R. Tice for NIEHS/NTP, R. Kavlock for the U.S. EPA, C. Austin for NCGC, and D. Jacobson-Kram for the U.S. FDA) that meets by conference call biweekly to discuss progress and resolve issue. There is also a designated co-chair for each working group from each agency; these scientists are jointly responsible for working group-specific activities. In addition, there is a general meeting of all Tox21 members held quarterly to discuss progress and future directions. All information connected with Tox21 is posted on a restricted website hosted by the U.S. EPA. Specific to the NTP, there is a NIEHS HTS Faculty chaired by Ms. Kristine Witt (BSB) that proposes, discusses, and makes recommendations in regard to HTS-related activities for NTP.

The Tox21 initiative is acquiring an international reputation for leadership in the introduction of innovative HTS technologies and computational approaches for identifying toxicity pathways and characterizing response to environmental exposures. Within the last year, representatives of Health Canada have attended one of our quarterly general meetings to discuss possible collaborations, representatives of the U.S. Department of Defense have become members of the Tox21 working groups, and representatives of the four Tox21 partners met with representatives of the European Commission Joint Research Centre, Institute for Health and Consumer Protection, to exchange information on our respective programs and to identify ways to exchange HTS assay protocols and information useful for the development of integrated methods for predicting compound toxicity.

To date, the Tox21 partners have published 35 articles and, as stated earlier, have presented almost 200 invited seminars and platform and poster presentations at national and international meetings to inform the research, testing, and regulatory communities of the purpose, goals, and progress related to our respective activities. The topics covered include:

- general information on Tox21 and its various components
- publicly accessible databases that are being developed or expanded to provide public access to Tox21-generated data as well as to the legacy data (i.e., data generated using standard toxicological tests) on the same compounds, if available
- results from qHTS/HTS primary screens and follow-up studies, as well as the results of chem- or bio-informatic analyses of the data generated using those assays
- informatic tools developed to evaluate data resulting from mid- and high-throughput *in vitro* and *in vivo* assays
- compound profiling for developmental, reproductive, chronic toxicity
- prioritization approaches for identifying compounds with, for example, endocrine activity for more extensive testing

The activities of Tox21 are helping to identify toxicity pathways, close data gaps, and suggest modes of action for toxic compounds. The knowledge gained will enable more efficient utilization of resources for the highest priority compounds. Also, as we gain more information on the biological effects of compounds across multiple targets, informatics and modeling efforts will provide more in-depth and quantitative molecular understanding of how biological systems respond to environmental compounds. These knowledgebases and corresponding *in silico* tools will reduce or quantify uncertainties relating to biological susceptibility, species differences, and dose response as part of a more efficient and intelligent targeted testing paradigm in support of testing prioritization for U.S Federal agencies.

I.4 The NTP Biomolecular Screening Branch

The goals of the Biomolecular Screening Branch (BSB) are to:

1) develop research and testing activities in high and medium throughput screening assays for rapid detection of biological activities of significance to toxicology and carcinogenesis

2) carry out the NTP's automated screening assays with *C. elegans*

3) develop analysis tools and approaches to allow an integrated assessment of HTS endpoints and associations with findings from traditional toxicology and cancer models

4) develop assays and approaches to understand the genetic and epigenetic bases for differences in susceptibility.

The BSB, when organized in 2007, included eight primary staff: four who were members of the *C. elegans* WormTox group, directed by Dr. Jonathan Freedman (Principal Investigator of the Comparative Genomics Group within the Toxicology and Pharmacology Branch of DIR) and four scientists with broad scientific experience. In response to the growth of our activities, the number of BSB staff has expanded also. In the last year, two scientists, two postdoctoral fellows, a part-time scientist from DERT, and two contractors with bioinformatic experience have been added to the branch. Also, NIEHS has made the decision to integrate the NTP Host Susceptibility Branch (HSB) into the BSB. The focus of the HSB is the genetic basis for differences in susceptibility to disease and adverse health effects that may lead to a better understanding of how substances in our environment may be hazardous to some individuals and not to others. By integrating the HSB into BSB, our ability to identify the key genes and pathways involved in a toxic response and the etiology of disease mediated by substances in our environment is greatly enhanced. An understanding of genes and environment interactions will lead to more specific and targeted research and testing strategies (both *in vitro* and *in vivo*) for NTP scientists to use for predicting the potential toxicity of substances in our environment and their presumptive risk to humans that may differ in disease susceptibility.

The BSB is functionally organized around five areas related to Tox21; the BSB is responsible also for the NTP genetic toxicity testing program. The four BSB subunits are:

- the Molecular Toxicology Group, which focuses on the identification and use of mid- and high-throughput screens for identifying biological effects potentially associated with human toxicity.

- the Pathway-Based Toxicology Group, which focuses on identifying toxicity pathways and their relation to diseases.

- the C. elegans Screening Core ("Worm Tox"), which develops and uses C. elegans mid-throughput assays as a complex model for evaluating environmental toxicants

- the Host Susceptibility Group, which focuses on the genetic basis for differences in susceptibility to environmental compounds and disease.

- the Statistical Analysis and Informatics Group, which focuses on the development and application of chem- and bio-informatic tools to characterize compound-specific responses in individual HTS or high content assays, as well as the response pattern(s) across multiple assays and their relationship to human (where available) and animal toxicological responses.

The data generated by the HTS program/ToxT1 Initiative is being use by NTP to help in the:

- identification/prioritization of compounds for more extensive toxicological testing

- ranking of compounds within the same class with regard to potential hazard

- categorization of multiple forms of a complex mixture (e.g., herbal products) into different "biological activity bins", based on patterns of response across multiple mid- and high-throughput screens

- identification of the kinds of animal studies that should be conducted and/or endpoints that should be evaluated in these studies, based on the pattern of activity of a compound across different mid- and high-throughput screens

- interpretation of results obtained in classical toxicological studies

- identification of the key cellular pathways linked to disease and the environmental compounds that might contribute to the appearance of that disease

I.5 Agenda for the BSB Review

The agenda for the BSC review is organized around three areas critical to the success of Tox21. First, representatives of the other Tox21 partners will inform the BSC about the roles of their organizations in the Tox21 Initiative. Please note that the contribution by the U.S. FDA to this report was delayed and will be provided at a later date. Second, an NTP representative from each Tox21 Working Group will present information about their group's activities. In addition, during this session, Dr. Ruili Huang (NCGC) will describe a new web-based informatics tool for the scientific community called "The Human Pathway Universe". Third, there will be a series of presentations related to specific activities that support the goals of the NTP HTS program and Tox21. These include a presentation on (1) the *C. elegans* "Worm Tox" Screening Facility, (2) probing mechanisms of inter-individual susceptibility to toxicants with population-based experimental approaches, (3) mining the NTP tissue archives for gene signatures, (4) a bioinformatics-based approach to identifying assays that query human health effects, and (5) the mouse methylome project. In addition, an afternoon poster session is planned for the first day, with posters keyed to these three different areas. The BSB review will conclude with my presentation "Future of Tox21 at NTP" where I will summarize our efforts to achieve the goals of the BSB and Tox21, present the main problems I think we face and how we might overcome those problems, and seek advice and recommendations from the BSC on how to improve on our activities.

1.6 Acknowledgements

There are many individuals who should be acknowledged for their significant contributions to the NTP HTS project and to Tox21. Our progress is due to the commitment of the BSB staff to achieving the Tox21 goals, the very active support of the NIEHS/NTP management, and the willingness of the Tox21 partners to mutually support a shared vision. In addition, we need to specially recognize Dr. Christopher Portier, while serving as the Associate Director for the NTP, for his vision in establishing and supporting the HTS program.

I.7 References

Collins FS, Gray GM, Bucher JR. 2008. Transforming environmental health protection. Science 319(5865):906-907.

II Tox21 Partners

This page left intentionally blank

II.1 Tox21 Partners:
U.S. EPA National Center for Computational Toxicology

This page left intentionally blank

II.1 Tox21 Partners: The U.S. EPA National Center for Computational Toxicology

David Dix, Ph.D.
Acting Deputy Director

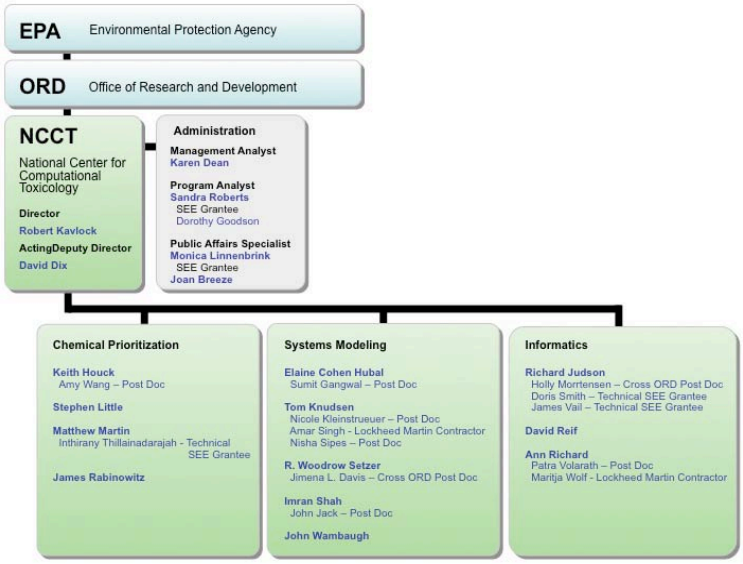
II.1.1 Background

The U.S. EPA's Computational Toxicology Research Program (CompTox) is a part of the Office of Research and Development (ORD). Located in Research Triangle Park, North Carolina,The CompTox Program:

- Coordinates and implements U.S. EPA's computational toxicology research program.
- Integrates innovative computing and information technologies with molecular biology to develop high-throughput decision support tools for assessing chemical exposure, hazard and risk.
- Coordinates research on chemical screening and prioritization, informatics and systems modeling.
- Collaborates with government and private organizations to leverage resources to further research capabilities.

CompTox is organized into four primary functional groups: (1) Chemical Prioritization, (2) Systems Modeling, (3) Informatics, and (4) Administrative. There are just over 20 permanent Federal staff in U.S. EPA's NCCT (U.S. EPA's center that coordinates the agency's CompTox research), federal staff from other ORD labs, collegiate-level academic institutions called the CompTox Science to Achieve Results (STAR) centers, a number of postdoctoral and predoctoral fellows, student contractors and Senior Environmental Enrollees (SEE)(**Figure II.1-1**).

Figure II.1-1 NCCT Organizational Chart



CompTox scientists work with other scientists in U.S. EPA program offices and regions, other federal agencies, industry, academia and stakeholder groups to revolutionize the current approach to assessing chemical toxicity risk to humans and the environment. A goal is to improve the capability of decision-support tools used to assess chemical risk to humans and the environment by:

- integrating advances in molecular biology, chemistry and computer science to more effectively and efficiently rank chemicals based on risk
- improving ongoing research projects and decision-support tools including Virtual Liver, Virtual Embryo, Toxicity Forecaster (ToxCast), Toxicity Reference Database (ToxRefDB), Aggregated data warehouse (ACToR) on over 500,000 chemicals, Exposure (ExpoCast™) and a public forum for publishing downloadable, structure-searchable standardized chemical structure files associated with toxicity data (DSSTox)

Brief information on the various NCCT programs are provided in **Figure II.1-2**.

Figure II.1-2 The U.S. EPA Computational Toxicology Research Program

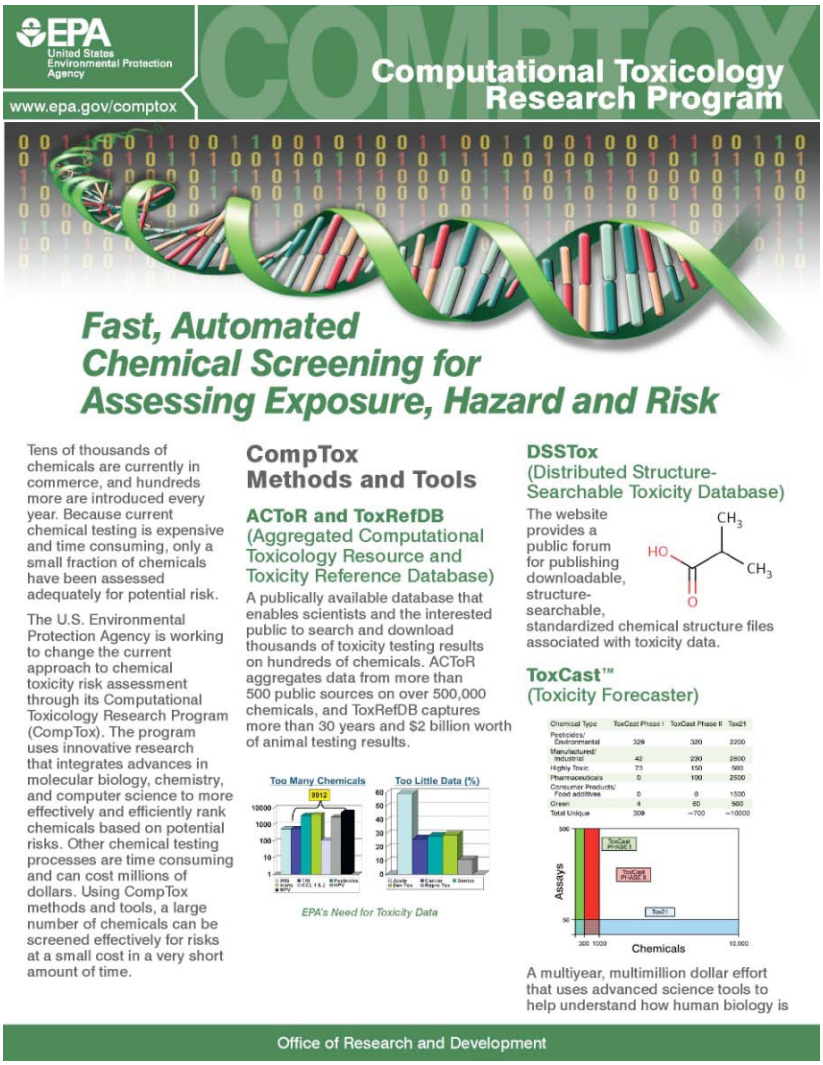


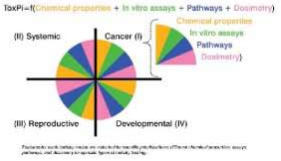
Figure II.1-2 The U.S. EPA Computational Toxicology Research Program (continued)

Computational Toxicology Research Program

impacted by exposure to chemicals and to determine which exposures are the most likely to lead to adverse health effects. ToxCast™ currently includes over 500 fast, automated chemical screening tests that are assessing more than 1000 environmental chemicals.

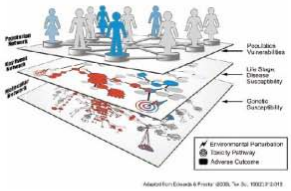
A large contributor to ToxCast™ is the Tox21 collaboration. Tox21 pools chemical research, data and screening tools from multiple federal agencies, including the Food and Drug Administration (FDA), the National Toxicology Program/ National Institute of Environmental Health Sciences and the National Human Genome Research Institute/ National Institutes of Health (NIH) Chemical Genomics Center.

ToxPi (Toxicological Priority Index)



A flexible prioritization support software tool that incorporates profiles, inferred toxicity pathways, dose estimates and chemical structural descriptors to calculate a potential toxicity score.

ExpoCast (Exposure Forecaster)



Systems Exposure Science: Extending Network Analysis

Works with ToxCast™ to determine priority chemicals that may need further screening and develops novel approaches for evaluating chemicals based on potential for biologically relevant human exposure.

Virtual Tissues

v - LIVER
VIRTUAL LIVER PROJECT

Advanced computer models of the liver and embryo that will be used to predict the effects of chemicals in humans. Virtual Liver (v-Liver™) uses fast, automated chemical screening data from ToxCast™ and other data to simulate how chemicals could cause liver disease and cancer in people. Virtual Embryo (v-Embryo™) uses ToxCast™ data to develop predictions for what chemical interactions will most likely lead to toxicity and birth defects.

Collaboration Opportunities

The CompTox Research Program partners and collaborates with EPA regions and program offices, industry, academia, trade associations, other federal agencies, state and local government agencies and non-governmental organizations with an interest in revolutionizing the current approach to assessing chemical toxicity risk to humans and the environment. Collaboration opportunities include a Communities of Practice group and different types of agreements that facilitate the sharing of research data and studies.

The CompTox program goal is to provide fast, automated tests for screening and assessing chemical exposure, hazard and risk. Housed within EPA's Office of Research and Development, CompTox is composed of three main elements. The largest component is the National Center for Computational Toxicology (NCCT), which was established in 2005 to coordinate research on chemical screening and prioritization, informatics and systems modeling.

The second element consists of research in EPA's National Health and Environmental Effects Research Laboratory (NHEERL) and National Exposure Research Laboratory (NERL). The final components are the academic centers working on various aspects of computational toxicology funded by EPA's Science to Achieve Results (STAR) program.

More information at:
www.epa.gov/comptox

National Center for Computational Toxicology

Robert Kavlock
NCCT Director
kavlock.robert@epa.gov

David Dix
NCCT Acting Deputy Director
dix.david@epa.gov

Monica Linnenbrink
NCCT Public Affairs Specialist
linnenbrink.monica@epa.gov

Main Office: 919-541-4219
www.epa.gov/comptox
109 T.W. Alexander Drive (B-205-01)
Research Triangle Park, NC 27711

II.1.2 ToxCast™: Predicting Hazard, Characterizing Toxicity Pathways, and Prioritizing the Toxicity Testing of Environmental Chemicals

In 2007, the U.S. EPA launched ToxCast™ to develop a cost-effective approach for efficiently prioritizing the toxicity testing of thousands of chemicals. ToxCast™

- Uses data from state-of-the-art HTS bioassays.
- Builds statistical and computational models to forecast potential chemical toxicity in humans.

- In 2007, Phase I provided U.S. EPA regulatory programs with science-based information helpful in prioritizing chemicals for more detailed toxicological evaluations and more efficient use of animal testing.

- Phase I profiled over 300 well-characterized chemicals (primarily pesticides) in over 400 HTS endpoints. Endpoints include biochemical assays of protein function, cell-based transcriptional reporter and gene expression, cell line and primary cell functional, and developmental endpoints in zebrafish embryos and embryonic stem cells.

- Phase 1 chemicals have already been tested using traditional toxicology methods including developmental toxicity, multi-generation reproductive studies, and sub-chronic and chronic rodent bioassays. ToxRefDB is the relational database storing this information- nearly $2 billion worth of animal toxicity studies.

- In 2010, Phase II of ToxCast will screen additional chemical compounds representing broader chemical structure and use classes to evaluate the predictive toxicity signatures developed in Phase I.

- Toxicity signatures from ToxCast will be defined and evaluated by how well they predict outcomes from mammalian toxicity tests and identify toxicity pathways relevant to human health effects.

- Provides the Tox21 collaboration access to ToxCast high-throughput screening assays and chemical library to increase the data available on the nearly 10,000 chemicals being studied.

II.1.3 CompTox Partners

The most important partnership to the U.S. EPA's CompTox research program is the Tox21 collaboration with the NIEHS/NTP, the NIH/NHGRI NCGC, and the U.S. FDA. Through a Memorandum of Understanding, our common purpose is to research, develop, validate and translate innovative chemical testing methods that characterize toxicity pathways. More recently, CompTox has entered into a partnership with the European Commission Joint Research Centre, Institute for Health and Consumer Protection (IHCP), to exchange research materials and results useful for the development of integrated methods for predicting chemical toxicity. U.S. EPA's CompTox research, through ToxCast™, is generating toxicological profiles of hundreds of reference chemicals using a comprehensive array of automated high throughput screening assays. ToxCast™ fits well with the work program of the IHCP's chemical safety area, where efforts are focused on the design and evaluation of integrated testing strategies for predicting chemical toxicity, by combining chemical-grouping approaches, computational modeling and *in vitro* testing

Other more recent CompTox partnerships include:

- Four pharmaceutical companies sharing failed drug products.

- Numerous academic institutions with ongoing computational toxicology research studies.

- Consumer products companies such as L'OREAL that are interested in using the CompTox tools to ensure safe products.

- International partners such as the European Commission's Registration, Evaluation, Authorization and Restriction of Chemicals project.

Organized by the U.S. EPA, any individual with an interest in promoting the usage of computational toxicology and exposure science, can join the Computational Toxicology and Exposure Science Communities of Practice.

Partnerships are finalized through numerous types of agreements including Cooperative Research and Development Agreements, Materials Transfer Agreements and Memorandum of Understanding.

II.1.4 Tox21 Related Publications

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95(1):5-12.

Houck KA, Dix DJ, Judson RS, Kavlock RJ, Yang J, Berg EL. 2009. Profiling bioactivity of the ToxCast chemical library using biomap primary human cell systems. J Biomol Screen 14(9):1054-66.

Houck KA, Kavlock RJ. 2008. Understanding mechanisms of toxicity: insights from drug discovery research. Toxicol Appl Pharmacol 227(2):163-78.

Judson R. 2010. Public Databases Supporting Computational Toxicology Journal of Toxicology Environmental Health - Part B: Critical Reviews 13(2):218-31.

Judson R, Elloumi F, Setzer RW, Li Z, Shah IA. 2008. A Comparison Of Machine Learning Algorithms For Chemical Toxicity Classification Using A Simulated Multi-Scale Data Model Bmc Bioinformatics Biomed Central Ltd, London, Uk, 9:241-256.

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118(4):485-92.

Judson R, Martin MT, Reif D, Houck KA, Knudsen TB, Rotroff D, et al. 2010. Analysis of eight oil spill dispersants using rapid, in vitro tests for endocrine and other biological activity. Environ Sci Technol 44(15):5979-5985.

Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, et al. 2009. The toxicity data landscape for environmental chemicals. Environ Health Perspect 117(5):685-95.

Kavlock RJ, Austin CP, Tice RR. 2009. Toxicity testing in the 21st Century: Implications for human health risk assessment. Risk Anal 29(4):485-487

Kavlock RJ, Dix DJ. 2010. Computational toxicology as implemented by the US EPA: Providing high throughput decision support tools for screening and assessing chemical exposure, hazard and risk. J Toxicol Environ. Hlth - Part B: Critical Reviews 1(2-4):197-217

Kavlock RJ, Reif DM, Richard AM, Rotroff DM, Romanov S, Medvedev A, et al. 2010. Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. Chem Res Toxicol 23(3):578-90.

Knight AW, Little S, Houck KA, Dix DJ, Judson R, Richard AM, et al. 2009. Evaluation Of High-Throughput Genotoxicity Assays Used In Profiling The US EPA Toxcast Chemicals. Regul Toxicol Pharmacol 55(2):188-99.

Martin MT, Dix DJ, Judson R, Kavlock RJ, Reif D, Richard AM, et al. 2010. Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. Chem Res Toxicol 23(3):578-90.

Martin MT, Judson R, Reif D, Kavlock RJ, Dix DJ. 2009. Profiling Chemicals Based On Chronic Toxicity Results From The US EPA Toxref Database. Environ Health Perspect 117(3):392-399.

Martin MT, Mendez E, Corum DG, Judson R, Kavlock RJ, Rotroff D, et al. 2009. Profiling The Reproductive Toxicity Of Chemicals From Multigeneration Studies In The Toxicity Reference Database. Toxicol Sci 110(1):181-90.

Reif D, Martin M, Tan S, Houck K, Judson R, Richard A, et al. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ Health Perspect, published 08 Sep 2010, doi:10.1289/ehp.1002180.

Rotroff D, Beam AL, Dix DJ, Farmer A, Freeman KM, Houck KA, et al. 2010. Xenobiotic-Metabolizing Enzyme And Transporter Gene Expression In Primary Cultures Of Human Hepatocytes Modulated By Toxcast Chemicals. J Toxicol Environ Health B Crit Rev 1(2-4):329-346.

II.2 Tox21 Partners: The NIH Chemical Genomics Center

This page left intentionally blank

II.2 Tox21 Partners: The NIH Chemical Genomics Center

Christopher P. Austin, M.D.
Director

II.2.1 Background

The NHGRI NCGC (www.ncgc.nih.gov; see **Figure II.2-1** for a simplified organizational chart) is an ultrahigh-throughput screening, informatics, and chemistry center that develops chemical probes of gene and cell functions, starting points for the development of new therapeutics for rare and neglected diseases, bioactivity profiles of large chemical and RNAi libraries, and new technologies to improve the efficiency and reach of small molecule technologies across the genome. Founded in 2004 and located within the NHGRI, the NCGC collaborates with over 100 researchers from academic, government, foundation, and biotechnology laboratories throughout the U.S. and the world. The NCGC's staff of 85 biologists, chemists, engineers, and informatics scientists have enormous breadth and depth of experience from the finest pharmaceutical, biotechnology, and academic organizations. Utilizing its titration-based qHTS paradigm and high-quality chemical libraries of over 400,000 compounds, the NCGC performs a new screen every week across >3 million wells, and has produced leads for a large number of disease targets, several of which are now in more advanced stages of drug development. As a founding member of the Tox21 Consortium with the NTP, the U.S. EPA, and the U.S. FDA, the NCGC is helping to transform toxicology assessment into a predictive, mechanistic, and efficient science. Over the longer term, the NCGC aims to define new relationships between proteins encoded by the human and other genomes via their shared interactions with small molecules. These insights will provide both a fundamentally new view of genome organization, and will assist in the development of small molecule compounds as new drugs to treat human disease.

II.2.2 NCGC Accomplishments in FY2010

During fiscal year 2010, the NCGC worked with over 300 researchers worldwide to advise them on assay design and development, chemistry research, informatics research, technology development projects, and to run high-throughput screens and chemically optimize small molecule leads. In collaboration with the Molecular Libraries Probe Production Centers Network (MLPCN), the U.S. EPA, the NTP, NIEHS, the U.S. FDA, National Cancer Institute (NCI), numerous rare disease foundations, and other intramural and extramural laboratories, the NCGC performed over 60 high-throughput screens on molecular targets and cellular phenotypes important for virtually every area of biology and disease. The NCGC also continued its work in the field of siRNA, completing 2 full-scale siRNA screening campaigns including follow-up, 3 primary screens (7,000 genes), and 6 pilot screens (1,000 genes). Twenty new chemical probes of diverse biologies were discovered, and NCGC scientists published 43 papers during FY10. The NCGC obtained 3 patents during this reporting period. Also during the reporting period, the NCGC deposited 195 BioAssays, 38 Summary AIDs, 7,734,000 concentration response curves, and a total of 46,404,000 data points into PubChem. NCGC continued to apply its chemistry expertise to optimizing probes; 28 chemistry projects were implemented during the year.

Under the joint leadership of the NCGC, the NIEHS/NTP, the U.S. EPA, and most recently the U.S. FDA, the Toxicology in the 21st Century project (Tox21) continued to flourish. Tox21 is an initiative designed to predict the toxicity of chemicals on human health and the environment. This is accomplished by

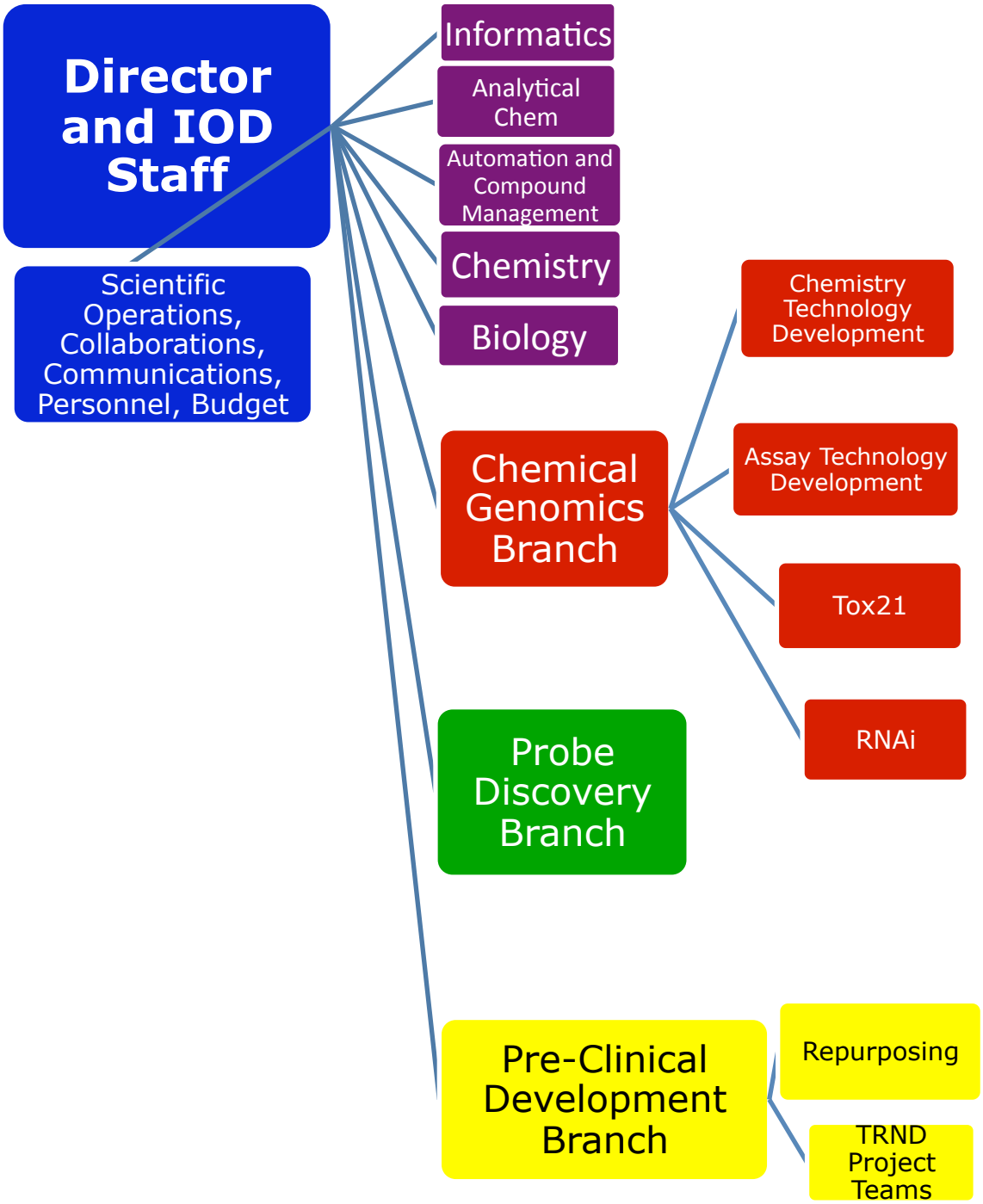


Figure II.2-1 NCGC Organizational Chart

developing in vitro assays for more predictive, mechanistically-based methods than those used with current animal testing. In July 2010, the U.S. FDA joined the collaboration; the U.S. FDA brings human toxicity data into the project, along with other expertise, to improve upon current chemical testing methods. In the current reporting year, NCGC completed 20 full Tox21 screens and 152 smaller-scale screens on specific, varied cell lines. Given the NCGC's continual efficiency improvement program, we were able to increase the throughput of existing robotic screening, informatics, and chemistry systems by improvements in applications, software, and utilization scheduling, driven by both the project teams and Project Management. The NCGC maintained its existing robotic technology and Bio Safety Levels 1, 2, and 3 facilities during the reporting period, in addition to adding a new robotic system dedicated to RNAi screening. The NCGC's Outreach program continued its extraordinary record of productivity during the reporting period. NCGC staff advised 245 outside investigators on assay design and assay development, and assisted over 40 investigators with chemistry, informatics, and technology development inquiries. NCGC scientists gave 104 invited presentations throughout the U.S., Europe, and Asia during the period. NCGC outreach resulted in the submission of over 60 assay development and screening applications for MLPCN programs. The NCGC website (ncgc.nih.gov) was maintained and is currently being completely redesigned, to be incorporated into an overall NCGC/TRND site. The Assay Guidance manual on the NCGC website has continued to evolve and has become a central resource for investigators interested in MLPCN science. To allow scientists across the world to share information on the topic, an Assay Wiki feature was designed and implemented during FY10. During the reporting period, the manual received 608,147 hits, with 48,437 unique visitors. 47% of these visitors originated from outside the US, from a total of 125 countries. The number of hits indicates that many Assay Guidance Manual readers are repeat visitors who find the resource useful to revisit on a frequent basis. During the year, the NCGC also maintained its status as an active member of the NCI's Chemical Biology Consortium. In addition, NCGC successfully competed for NIH funds dedicated to an induced pluripotent stem cell project, on which work has commenced. NCGC continued to work on its NIH Directors Challenge Award for malaria research. Finally, NCGCs work on its two IATAP grants continued and was submitted for renewal; the two IATAP awards include a program to modify non-nucleoside reverse transcriptase inhibitors to extend their potency range into newly characterized RT mutants, and a program to modify the natural product manicol to inhibit the RNAse domain of HIV RT.

II.2.3 NCGC Assay and Screening Capabilities

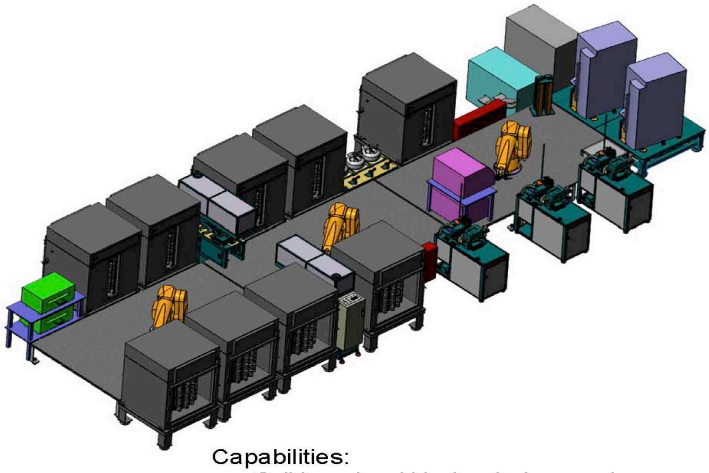
NCGC assay and screening capabilities are provided in the following table and figures. **Table II.2-1** provides information on the assay readouts, while **Figures II.2.-2** through **II.2.-6** provide a schematic view of the robotics facilities, including the new facility being built specifically for Tox21, and information on liquid handlers and plate readers.

Table II.2-1 Assay readouts at the NCGC

ASSAY FORMATS	SCREENING CAPABILITIES	BIOLOGICIAL EXPERTISE
Absorbance	1536 well plates; HTS (Primary and secondary screens)	Enzyme assays (kinase, phosphatase, protease, etc.)
AlphaScreen	1536 well plates; HTS (Primary and secondary screens)	Protein-protein interaction, replacement of ELISA assay, protein-peptide interaction
Fluorescence Intensity	1536 well plates; HTS (Primary and secondary screens)	Enzyme assays (kinase, phosphatase, protease, beta-lactamase reporter, etc.); Receptor binding assay
Fluorescence Polarization	1536 well plates; HTS (Primary and secondary screens)	Protein-peptide interaction, protein-DNA/RNA interaction, Kinase assay (IMAP)
Time Resolved Fluorescence	1536 well plates; HTS (Primary and secondary screens)	Redox enzyme systems
HTRF/LANCE (FRET)	1536 well plates; HTS (Primary and secondary screens)	cAMP assay, Kinase assay
Luminescence	1536 well plates; HTS (Primary and secondary screens)	Enzyme assays, luciferase reporter-gene assay, cytotoxicity/cell growth assay (ATP content)
Laser scanning cytometry (Acumen Explorer)	1536 well plates; HTS (Primary and secondary screens)	Low-resolution image-based high content assays
Microscopy-based imaging (GE INCell 1000) (MDS ImageExpress Micro)	96-, 384-, 1536 well plates; secondary screens	High-resolution image-based high content assays
FDSS-7000 Kinetic Reader	1536 well plates; HTS (Primary and secondary screens)	Intracellular calcium kinetic assay (GPCRs and Calicum channels), ion flex assay (Ion Channels), aequorin assay (GPCRs)
Real-time cell analyzer (ACEA impedance-based platform)	16 well strip/96 well plate; secondary kinetic assay	Cell growth rate, cytotoxicity kinetic measurement

Figure II.2-2

NCGC Screening System 1: BSL1/Kalypsys



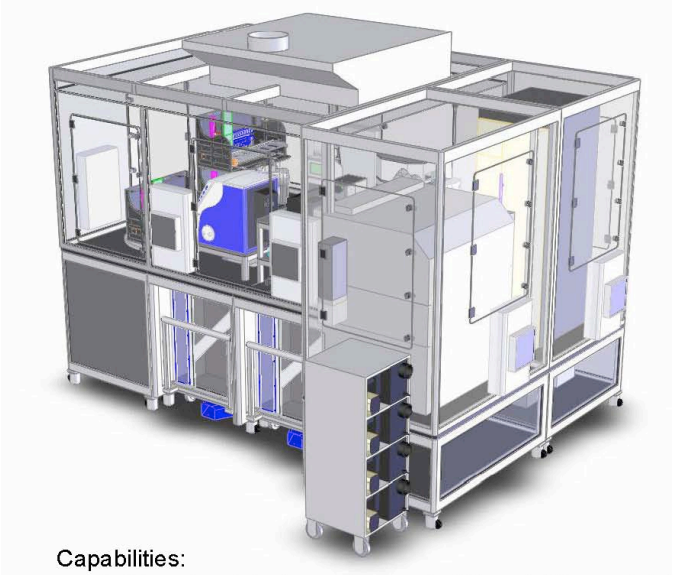
Capacity	• 3.0M Assay Wells • 5.0M Compound Wells
Throughput	1400 plates/2.2M wells/day (1536 format)
Readers	• ViewLux • Acumen (2) • Envision (2) • INCell1000
Features	• Integrated system with compound store, liquid dispensers, pin-tool transfer, centrifuge, incubators, and detectors • Custom operating software • Automatic loading and unloading stations • Dispense inspection using integrated CCD cameras

Capabilities:

- Cell-based and biochemical assays in 1536-well plate format
- Incubations up to 3 days
- Routinely runs new assay of >350,000 compounds at 7 concentrations each week

Figure II.2-3

NCGC Screening System 2: BSL2/HRE



Capacity:	•370K Assay Wells •370K Compound Wells
Throughput:	•~240 plates/day •(370K wells/day 1536) •(100K wells/day 384)
Readers:	•ViewLux •Acumen
Features:	- Modular approach to HTS - Three docking stations to quickly facilitate changing of both compound and assay plate storage - Flexible scheduling software to allow for complex assay methods - BSL2 rated

Capabilities:

- Cell-based and biochemical assays in 384- or 1536-well plate format
- Incubations up to 3 days 1536, 7 days in 384
- Routinely runs new assay of >100,000 compounds at 7 concentrations each week

Figure II.2-4

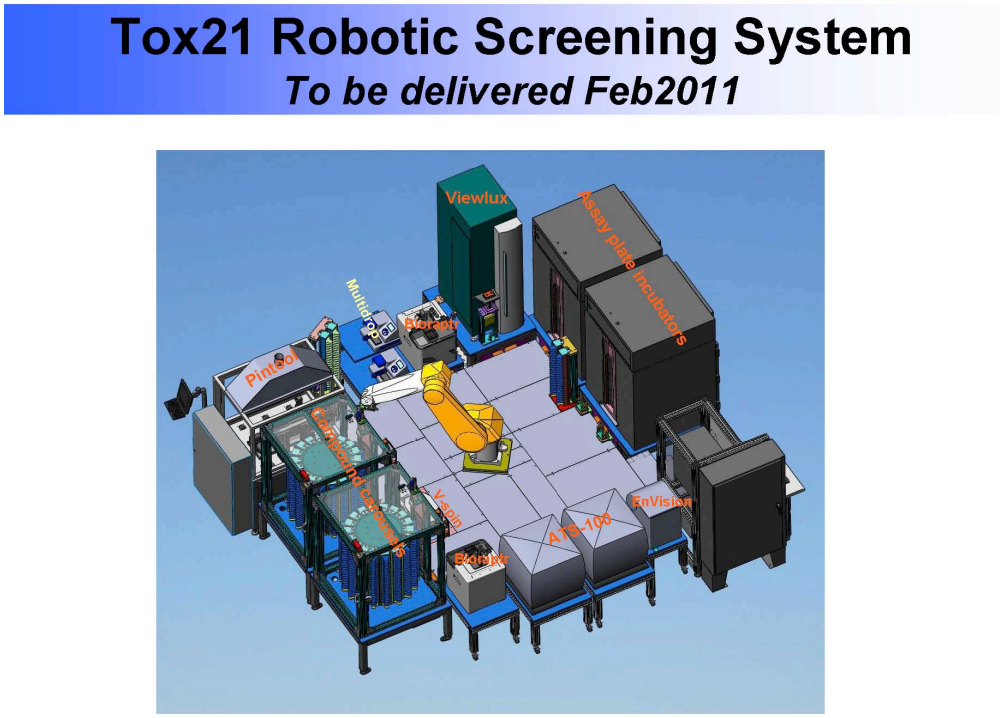


Figure II.2-5

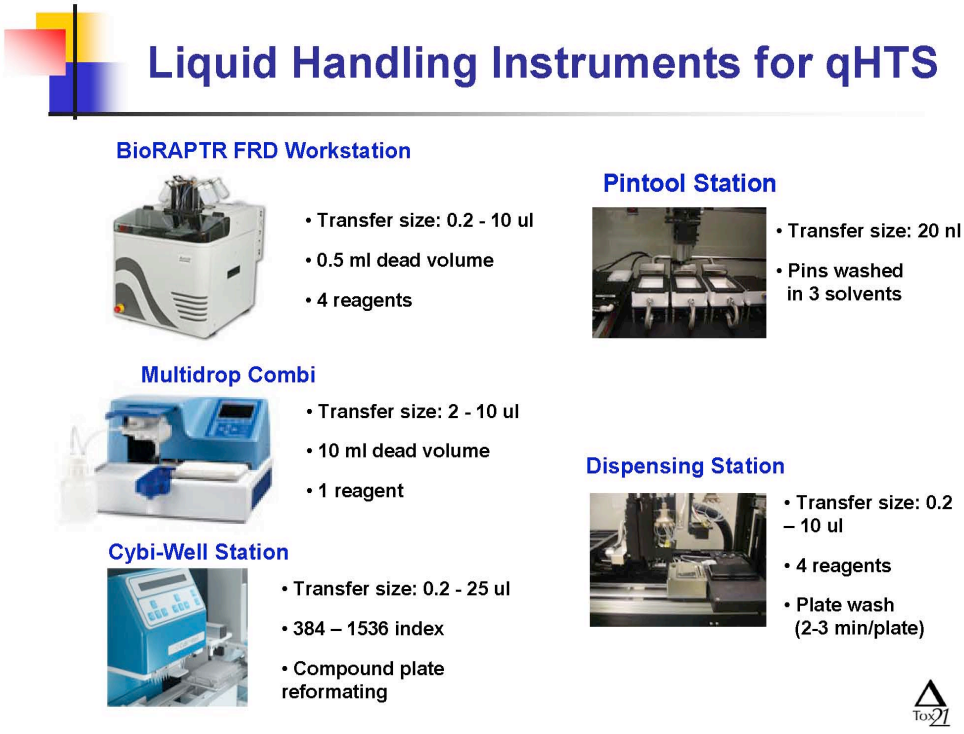
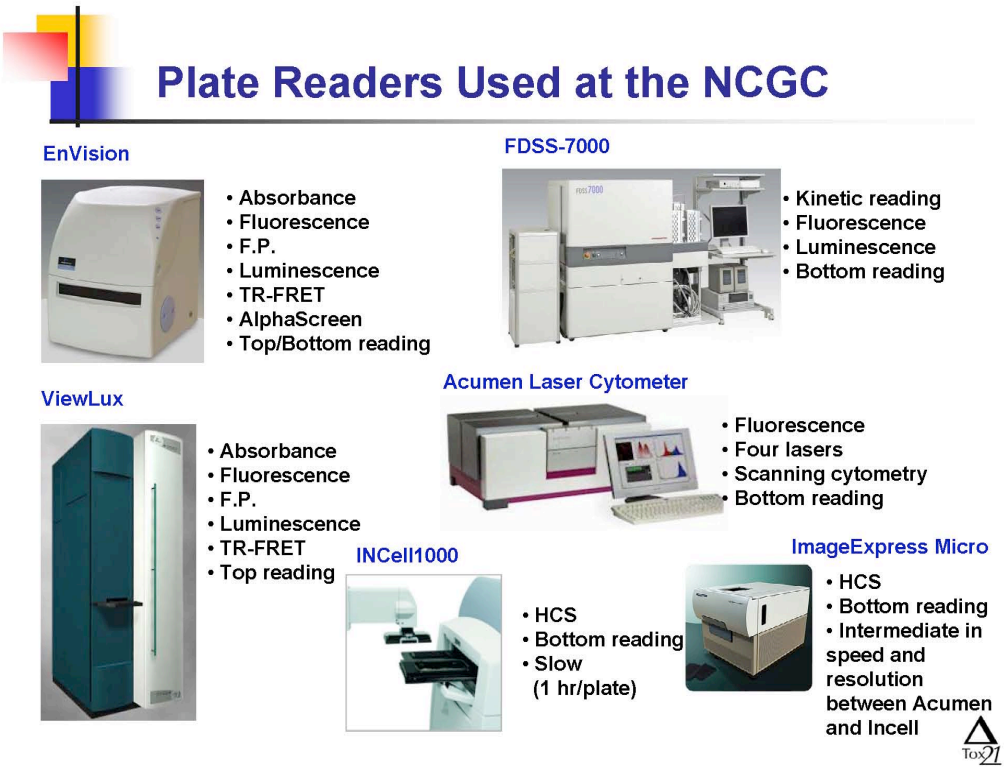


Figure II.2-6



II.2.4 Tox21 Related Publications

Cho M, Niles A, Huang R, Inglese J, Austin CP, Riss T, et al. 2008. A bioluminescent cytotoxicity assay for assessment of membrane integrity using a proteolytic biomarker. Toxicology In Vitro 22:1099-1106.

Huang R, Xia M, Cho M, Sakamuru S, Shinn P, K Houck, et al. 2010. Chemical genomics profiling of environmental chemicals on human nuclear receptors. Environ Health Perspect, submitted.

Huang R, Southall N, Cho M, Xia M, Inglese J, Austin CP. 2008. Characterization of diversity in toxicity mechanism using in vitro cytotoxicity assays in quantitative high throughput screening. Chem Res Toxicol 21:659-667.

Huang R, Southall N, Xia M, Cho M, Jadhav A, Nguyen D-T, et al. 2009. Weighted feature significance: a simple, interpretable model of compound toxicity based on the statistical enrichment of structural features. Toxicol Sci 1122:385-93.

Judson RS, Martin MT, Reif DM, Houck KA, Knudsen TB, Rotroff DM, et al. Analysis of Eight Oil Spill Dispersants Using Rapid, In Vitro Tests for Endocrine and Other Biological Activity. Environ Sci Technol 4415:5979-85.

Shukla SJ, Huang R, Austin CP, Xia M. In press. The future of toxicity testing: a focus on in vitro methods using a quantitative high throughput screening platform. Drug Discovery Today.

Titus S, Beacham D, Shahane S, Xia M, Huang R, Southall N, et al. 2009. A new homogeneous high throughput screening assay for profiling compound activity on the hERG channel. Anal Biochem 394:30-38.

Xia M, Huang R, Sun Y, Semenza GL, Aldred SF, Witt K, et al. In press. Identification of chemical compounds that induce HIF-1α activity. Toxicol Sci 112(1):153-163.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho M, et al. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect 116:284-291.

II.2.5 NCGC Infrastructure Publications

Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, et al. 2006. Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U.S.A. 103:11473-11478.

Inglese J, Johnson RL, Simeonov A, Xia M, Zheng W, Austin CP, et al. 2007. High-throughput screening assays for the identification of chemical probes. Nat Chem Biol 3:466-479.

Michael S, Auld D, Klumpp C, Jadhav A, Zheng W, Thorne N, et al. 2008. A robotic platform for quantitative high-throughput screening. Assay Drug Dev Technol 6 5:637-658.

III.1 Introduction to the Tox21 Working Groups

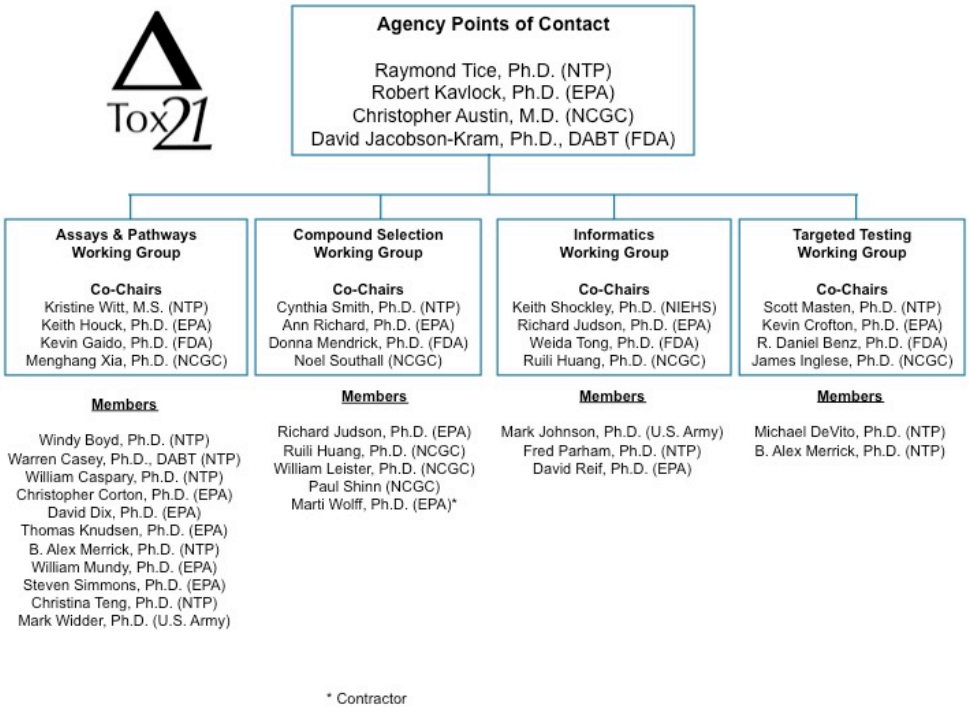
This page left intentionally blank

III.1 Introduction to the Tox21 Working Groups

Raymond Tice, Ph.D.
Chief, Biomolecular Screening Branch

Critical to the success of the Tox21 Collaboration are the four Working Groups - Chemical Selection, Pathways and Assays, Informatics, and Targeted Testing (**Figure III.1-1**). Each Working Group is co-chaired by a representative of each of the Tox21 partners; it is their responsibility to coordinate Working Group specific activities across agencies and to act as conduits for information between the agency points of contact and the members of each Working Group.

Figure III.1-1 Organization of Tox21



A brief summary of the activities of each Tox21 Working Group follows; more extensive information is provided in each Working Group report.

III.1.1 Chemical Selection Working Group

Developing a comprehensive, characterized compound library is critical to the ability of Tox21 to evaluate and compare the *in vitro* toxicity of compounds. Ideally, the identity, purity, concentration, and stability of compounds in this library should be known. The current primary focus of this Working Group is to establish a library >10,000 (10K) dimethyl sulfoxide (DMSO) soluble compounds with known structures for screening at the NCGC in Phase II, as well as a procedure for uniquely identifying and

tracking each compound; and to establish procedures for determining the identity, purity, and stability of each compound. Once completed, the focus of this Working Group will move to establishing a library of water-soluble compounds and one of mixtures for testing at the NCGC.

III.1.2 Assays and Pathways Working Group

The Tox21 Pathways and Assay Working Group has an extensive charge, central to the Tox21 screening efforts at the NCGC. First, this group has primary responsibility for identifying appropriate targets for screening environmental compounds, based on their biological significance to toxicological endpoints. Targets may include enzyme activity, gene regulation, pathway activation, changes in cell physiology, and biochemical reactions, for example. Second, this group identifies appropriate assays to screen the desired targets. Appropriate assays are those that have technically feasible assay protocols (compatible with NCGC experimental protocol requirements) and produce high quality data for use in toxicity profiling. Third, this group tracks assay activities, reviews assay data, and evaluates assay performance based on the quality of the data, and attempts to resolve problems that may arise with assays. Finally, this group conducts outreach to the international HTS community to stimulate collaborations in assay design and assay acquisition, and outreach to vendors and academic researchers to promote research and development into new assays or assay formats useful for toxicity profiling. Currently, the focus of this group has been on developing a strategy for screening the >10K library in Phase II at the NCGC and on identifying and evaluating methods for incorporating xenobiotic metabolism into the *in vitro* assays used to provide data for Tox21.

III.1.3 Informatics Working Group

The goals of the Tox21 collaboration are to identify mechanisms of toxicity, prioritize chemicals for *in vivo* testing, and predict harmful responses to environmental chemicals. The qHTS assays used at the NCGC play an important role in efforts to achieve these objectives, as the ability of a substance to induce a toxicological response is better understood by analyzing the response profile over a broad range of concentrations. The goals of this Working Group include developing (1) statistical for distinguishing between active, inactive, and inconclusive responses, and (2) informatic tools for evaluating the results obtained from testing conducted in support of Tox21 for predictive toxicity patterns. Another, critically important goal is to make all data generated in support of Tox21 publicly accessible via NTP's CEBS database, EPA's ACToR, and the National Center for Biotechnology Information's PubChem to encourage independent evaluations and/or analyses of the Tox21 test results.

III.1.4 Targeted Testing Working Group

As HTS data on compounds with inadequate testing for toxicity becomes available via Tox21, there will be a need to test selected compounds in more comprehensive assays. This tasks assigned to this Working Group are to (1) prioritize substances for more complex testing, including the use of alternative assay platforms or species (e.g., *C. elegans*, zebrafish embryos, rodents); and (2) identify and implement strategies for evaluating the relevance of prioritization schemes and prediction models developed by the Tox21 partners. Currently, through resources provided by the NTP, the Working Group is evaluating the *in vivo* relevance of a statistical model that uses ToxCast™ Phase I data to predict nongenotoxic rodent liver tumorigens.

III.2 Tox21 Chemical Selection Working Group

This page left intentionally blank

III.2 Tox21 Chemical Selection Working Group

Cynthia S. Smith, Ph.D.
Chief, Program Operations Branch

Tox21 Chemical Selection Working Group

Richard Judson, Ph.D. (U.S. EPA NCCT)
Ruili Huang, Ph.D. (NHGRI NCGC)
Donna Mendrick (U.S. FDA NCTR)*
Ann Richard, Ph.D. (U.S. EPA NCCT)*
Noel Southall, Ph.D. (NHGRI NCGC)*
Paul Shinn (NHGRI NCGC)
Cynthia Smith, Ph.D. (NTP POB)*
Marti Wolf, Ph.D. (Lockheed Martin)
William Leister (NHGRI NCGC)
** = Co-Chairs*

NTP Scientists

Bradley J. Collins, M.S.P.H. (POB)
Beby Jayaram, Ph.D. (POB)

III.2.1 Introduction

The establishment of a large, well-documented compound library is critical to the ability of Tox21 to evaluate, analyze, and compare the *in vitro* toxicity of compounds. Ideally, this library should be populated with materials of known identity, purity, and concentration and they should be compatible with the primary buffer system used in quantitative high throughput screening (qHTS) assays at the NCGC.

At the start of the HTS project, in 2006, the NTP and NCGC collaborated on a "proof of principal" with a small library of 1408 substances and a small number of assays compatible with 1536-well qHTS. The U.S. EPA's NCCT decided to join in these efforts by arranging for the NCGC to screen a second compound library of 1408 compounds, which was subsequently expanded to 1462 substances to ensure that all compounds included in ToxCast™ Phase I were also screened for activity in qHTS at the NCGC. The first NTP 1408 library contained 1353 unique compounds and 55 compounds in duplicate to assess assay reproducibility. Among the 1353 unique compounds, 1206 compounds that had been tested by the NTP in one or more standard *in vitro* (predominantly for genetic toxicity in the Salmonella reverse mutational assay) and/or *in vivo* (e.g., carcinogenicity, development toxicity, reproductive toxicity, immunotoxicity) tests. The NTP library also included 147 reference compounds identified by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) for the development of alternative test methods for dermal corrosion, acute toxicity, and endocrine activity. By use category, the NTP library included industrial compounds, pharmaceuticals, dye components, pesticides, natural products, and food additives, as well as organic and inorganic pollutants. The molecular weight (MW) of these compounds ranged from 32 to 1168, and their calculated log p from -3 to 13.2. The corresponding

U.S. EPA library of 1462 substances included 1384 unique compounds and 78 duplicates; MW ranged from 58 to 516 and calculated log p from -2.8 to 8.2. There was about a 400 compound overlap between the two libraries.

The NCGC qHTS system uses a 1536-well plate with the first 4 columns used for internal quality control (QC). The pin tools for distributing the test compounds are set currently to use DMSO as the solvent for stock formulations and as a diluent for the preparation of "daughter" plates, which are serial dilutions of the stock formulations. Therefore, the compound library had to be composed of materials soluble in DMSO. The stock formulations were set at 10 mM for the NTP library and 20 mM for the later U.S. EPA library. Based on their standard dilution scheme, this meant that cell-based assays would typically be testing compounds at a maximum concentration of ~50 and ~100 μM for the NTP and U.S. EPA libraries, respectively. To screen both libraries up to the same concentration, the ~50 μM dilution of the NTP library was added twice to each well to prepare the ~100 μM plate.

These compounds were not analyzed for purity, identity, or concentration in DMSO prior to their use at the NCGC. During examination of the results generated with these pilot libraries, a few discrepancies and issues were identified that would direct efforts to design future libraries. We learned from this effort that stability in the solvent is an issue for some compounds, that results from multiple lots from multiple sources of the same compound are more easily interpreted if the purity and identity of each analyte is unequivocally established before use, and that a significant amount of information from certificates of analysis accompanying purchased chemicals is inaccurate.

III.2.2 Current Efforts

Once our initial efforts to use HTS techniques to screen the first compound libraries from the NTP and U.S. EPA for potentially toxicological relevant effects were successful, other researchers interested in screening environmental compounds for biological effects began to inquire about using some of the library compounds in their research. The NTP decided to make the first 1408 compounds available through the Molecular Libraries Screening Committee Network repository (see http://mli.nih.gov/mli/). This repository is housed at DPI-BioFocus in San Francisco, CA, and contains chiefly pharmaceutical compounds that are used to screen for drug reactivity. DPI-BioFocus declined to handle approximately half of the NTP library chiefly because of their various chemical handling hazards, but requested approximately 700 of the NTP compounds. To date, ~600 compounds have been submitted to DPI-BioFocus, one was rejected for QC reasons. An additional 100 are in the process of being characterized before submission to the library.

With the success of many aspects of this first "proof of principal", an agreement (the 2008 Memorandum of Understanding described in Section I) was signed by the NTP, the NCGC, and the U.S. EPA to formally collaborate on the research, development, validation, and translation of new and innovative test methods that characterize key steps in toxicity pathways. Also formalized in this agreement was the establishment of the Chemical Selection Working Group (CSWG), among several others. This group meets on a regular basis to coordinate chemical selection activities originally across the three agencies, and now with the U.S. FDA as well. In this way, compound selection, formulation, and tracking strategies are developed and kept uniform so that the end result will be a compound library with components submitted by different organizations, but that has uniform supporting information for each compound regardless of origin. With this structure in place, the CSWG began planning for a large library with a goal of 10,000 compounds. Approximately 4000 compounds, predominantly drugs, were to be obtained by NCGC and handled in their facility. For environmental compounds, NTP and U.S. EPA collaborated on a database to contain every compound appearing on a list evaluating toxicity as well as

list of compounds of interest because of potential toxicity or because of structural relationships with known toxins. This list of approximately 120,000 compounds was then culled for duplicates, compounds with no known structure, and mixtures. Once these compounds had been excluded, ~11,000 compound remained that were considered candidates for screening at the NCGC. Subsequently, an estimated 2000 more compound nominations have been included and considered for inclusion in the "10K" library. U.S. EPA has selected approximately 4300 compounds to evaluate for inclusion in the library, all of the compounds that the U.S. EPA plans to include in ToxCast™ Phase II will also be included in the 10K library designated for qHTS screening at the NCGC, and NTP is working from a list of approximately 4000 compounds. Of special interest is a set of ~100 pharmaceuticals that failed in clinical trials, and that have been provided to the U.S. EPA through Material Transfer Agreements by different pharmaceutical companies for inclusion in ToxCast™ Phase II and the Tox21 10K library, with the agreement that the identity, structure, and toxicity data would be made public.

The next steps included purchasing and formulating the compounds in DMSO. NCGC conducted these activities in their own laboratory, while NTP and U.S. EPA did so using contracted resources. For NTP, the chemical support contractor not only purchased the compounds and formulated them, but also performed a physical/chemical properties screen prior to ordering and a screening purity and identity confirmation on the bulk materials. The property cutoffs were: a MW range of between 100 and 1000, a vapor pressure less than 10 Pa, and a calculated log P value of -2 to 6. A number of nominated compounds were not found to be commercially available, some number more were not available in acceptable purity (95 % or better), and still others were not soluble in DMSO in sufficient concentration (20 mM is the desired stock concentration although some compounds of special interest are being included in the library at lower stock concentrations in DMSO). In these cases, the compounds were flagged in the inventory and not included in the library. Regular meetings of the CSWG include discussion of progress with obtaining acceptable compounds and coordination of the preparation of the plates for submission to the NCGC.

An issue of special interest to the Tox21 partners is assay reproducibility. Thus, it was determined that a designated set of duplicate compounds should be included in each plate. U.S. EPA scientists led this effort by evaluating the list of candidate compounds identified as active in different assays in Tox21 Phase 1 and identified a representative and statistically meaningful set of 88 compounds that cover chemical space. The CSWG agreed that variability in these compounds should be minimized, so U.S. EPA agreed to ask their contractor to procure the selected compounds and formulate them for all three participating organizations. That effort is in progress, but when completed, the replicates will appear on each 1536-well assay plate in a manner designed to account for any affects in the results due to placement of a chemical on the plate (border effects, etc).

In the end, the total library will be approximately 11,000 diverse compounds in classes that range across industrial chemicals, sunscreen additives, flame retardants, pesticide additives and their metabolites, plasticizers, solvents, food additives, natural product components, drinking water disinfection byproducts, preservatives, therapeutic agents, and synthesis byproducts. As this is a work in progress, a listing of compound use types, molecular weight range, and other descriptions of the chemical space will be deferred until the list is complete.

To address issues of compound purity and identity, QC analyses of the entire library is being conducted under contract using a tiered approach. First, a high-throughput, high performance liquid chromatography system with multiple detectors (mass spectrometry, ultraviolet diode array, evaporative light scattering [ELSD], and chemiluminescent nitrogen detection [CLND]) is being used to perform identity characterization and purity estimation. These analyses are conducted on 1536-well

plates prepared by NCGC for QC purposes at the same time the assay plates are prepared. Identity confirmation is performed by a matching molecular ion in the mass spectrum with the desired compound; purity analysis is conducted with the ELSD. For compounds containing nitrogen, the CLND lends a measure of quantitation of the compound concentration in the sample. This system works well for many of the compounds in the library, but will not generate useful data for some of the more volatile compounds and those that will not properly ionize in the mass spectrometer. An agreement has been reached with the National Institute for Standards and Technology (NIST) to conduct follow-up analyses by gas chromatography with mass spectrometry and other techniques as needed. Initial QC analyses are currently underway using plates submitted by NCGC and preliminary results indicate that this approach will provide identity and purity information on most, if not the entire library and will estimate concentration of many library entries. Chromatographic and QC data for all components of the Tox21 Library will be available to support assay results.

III.2.3 Future Directions

Two additional issues remain for future work by this group. Quantitation of the compounds in DMSO is a complex and difficult task given the diversity of the chemical properties compounds to be included in the collection. At this point, we have not found a cost-effective approach to confirming concentration for each compound in the library. Further, it is desirable to quantitate the compounds in the actual test well, usually a buffered aqueous system, to ensure that the material is available as a challenge to the test system. This adds a requirement for additional sensitivity, since the plates are assayed at far lower concentrations than the “stock” plate, and the analytical system has to be compatible with water and buffers. These are challenges that are not met with current technology in a high-throughput and cost effective manner, but they remain goals for the CSWG.

Additional future plans for the CSWG include the design of plates using solvent systems that are appropriate for compounds not sufficiently soluble in DMSO. These solvent systems will restricted to those compatible with cell-based assays, so the likely first attempt will be to construct a plate with water as a solvent to include hydrophilic compounds that were not sufficiently soluble in DMSO. This will expand the library of compounds by at least 10 %. The search for a less polar solvent to bring in additional compounds, yet be compatible with the assays is underway. In addition, Tox21 is interested in establishing a “mixtures” library in order to evaluate interactions between compounds in the qHTS assays used at the NCGC.

III.2.4 References

Janzen WP, Popa-Burke, IG. 2009. Advances in Improving the Quality and Flexibility of Compound Management. J Biomolecular Screening 14(5):444-451.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. 2007. Compound Cytotoxicity Profiling Using Quantitative High-Throughput Screening. Environ Health Perspect 116(3):284-291.

III.3 Tox21 Assays and Pathways Working Group

This page left intentionally blank

III.3 Tox21 Assays and Pathways Working Group

Kristine Witt, M.S.
Biomolecular Screening Branch

Tox21 Assay and Pathways Working Group

Windy Boyd, Ph.D. (NTP BSB)
Warren Casey, Ph.D., DABT (NICEATM)
William Caspary, Ph.D. (NTP BSB)
Chris Corton, Ph.D. (U.S. EPA NHEERL)
David Dix, Ph.D. (U.S. EPA NCCT)
Kevin Gaido, Ph.D. (U.S. FDA CVM)*
Keith Houck, Ph.D. (U.S. EPA NCCT)*
Thomas Knudsen, Ph.D. (U.S. EPA NCCT)
B. Alex Merrick, Ph.D. (NTP BSB)
William Mundy, Ph.D. (U.S. EPA NHEERL)
Steve Simmons, Ph.D. (U.S. EPA NHEERL)
Christina Teng, Ph.D. (NTP BSB)
Mark Widder, Ph.D. (U.S. DOD U.S. Army)
Kristine Witt, M.S. (NTP BSB)*
Menghang Xia, Ph.D. (NHGRI NCGC)*

** = Co-Chairs*

NIEHS/NTP High Throughput Screening Faculty

Scott Auerbach, Ph.D., DABT (BSB)
Windy Boyd, Ph.D. (BSB)
Chad Blystone, Ph.D. (TB)
Warren Casey, Ph.D., DABT (NICEATM)
Bradley Collins, M.S.P.H. (POB)
Michael Cunningham, Ph.D., DABT (BSB)
Michael DeVito, Ph.D. (TB)
John French, Ph.D. (BSB)
Michelle Hooth, Ph.D. (TB)
Scott Masten, Ph.D. (NTP Office of Nomination and Selection)
Barry McIntyre, Ph.D., DABT (TB)
Michael Sanders, Ph.D. (POB)
William Schrader, Ph.D. (Office of the Scientific Director)
Daniel Shaughnessy, Ph.D. (DERT)
Keith Shockley, Ph.D. (BB)
William Stokes, D.V.M., D.A.C.L.A.M. (NICEATM)

Associated NIEHS Scientists

Trevor Archer, Ph.D. (LMC)
Perry Blackshear, M.D., D.Phil. (LST)
Anton Jetten, Ph.D. (LRB)
Masahiko Negishi, Ph.D. (LRDT)
Samuel Wilson, M.D. (LSB)

Collaborating Scientist

Ruili Huang, Ph.D. (NHGRI NCGC)
Richard Judson, Ph.D. (U.S. EPA NCCT)
Thomas Knudsen, Ph.D. (U.S. EPA NCCT)
William Mundy, Ph.D. (U.S. EPA NHEERL)
Matias Attene Ramos, Ph.D. (NHGRI NCGC)
Ivan Rusyn, M.D., Ph.D. (Univ. North Carolina, Chapel Hill)
Sunita Shukla, Ph.D., (NHGRI NCGC)
Anton Simeonov, Ph.D. (NHGRI NCGC)
Kimiyo Yamamoto (Public Health and Hygiene, Graduate School of Medicine, Osaka Medical College, Osaka, Japan)
Shunichi Takeda, Ph.D. (Department of Radiation Genetics, Kyoto University Medical School, Kyoto, Japan)

III.3.1 Background

III.3.1.1 Major responsibilities of the Tox21 Assays Working Group

1. To identify appropriate targets for screening extensive compound libraries, based on biological significance to toxicological endpoints. Targets include enzymes, gene regulation, pathway activation, changes in cell physiology, and biochemical reactions.

2. To identify appropriate assays to screen the desired targets. Appropriate assays are those that are technically feasible and produce high quality data for use in toxicity profiling.

3. To track assay activities, review assay data, and evaluate assay performance based on the quality of the data, and resolve problems that may arise with assays.

4. To conduct outreach to the international HTS community to stimulate collaborations in assay design and assay acquisition, and to outreach to vendors and academic researchers to promote research and development into new assays or assay formats useful for toxicity profiling.

III.3.2 Overview

Our Tox21 Phase I screening activities over the past five years have provided a unique experience in the use of HTS approaches for the toxicity screening of large libraries of environmental compounds[1]. The knowledge gained has greatly enhanced our understanding of how HTS can be effectively used for the toxicity profiling of compounds for the purpose of prioritization for more comprehensive toxicological testing, for understanding mechanisms of action related to the toxic effects of compounds, and potentially for predicting human health effects. Traditional toxicity testing approaches are not capable of evaluating the tens of thousands of compounds to which humans are exposed for hazard. HTS has been extensively used for drug discovery; drug discovery aims to identify small molecules that are active at physiologically achievable concentrations and that interrupt an adverse cellular process involved in a disease process with no additional liabilities. The goal of toxicological screening is to identify any compound that can disrupt normal cellular processes, potentially causing physiological changes that eventually may result in cell death or initiation of a disease process (Golke et al. 2009).

Our screening partner, the NCGC, exploits a recently developed technology suitable for toxicological investigations: qHTS (Inglese et al. 2006). With qHTS, tens of thousands of compounds can be rapidly screened at multiple concentrations (typically 7-15) to yield detailed concentration-response curves defining compound activity in a biologically relevant assay at physiologically appropriate concentration levels. The promise of this approach for toxicological evaluations is considerable but limitations have also been identified. Limitations on assay

[1] "Environmental compounds" is the term used to describe our Tox21 compound collection that includes industrial compounds, pharmaceutical compounds, environmental compounds, and other compounds of interest to Tox21; the term is used to distinguish a library such as ours from the more typical small molecule library used in drug discovery screening.

protocols are imposed by the use of 1536-well plates on a highly automated robotics platform (**Table III.3-1**). We have been constrained in our assay selection by currently available technologies, but we have been able to adapt assays to better conform to our technological requirements. Through SBIRs, research collaborations, and communication with commercial assay suppliers, we have been successful in promoting interest in the development of HTS assays compatible with our requirements.

Table III.3-1. NCGC criteria for acceptable qHTS assays

Criteria	Biochemical	Cell-based
*Plate Format**	96-well or higher. 1536-well format is preferred. Assay volume 2-6 µL	96-well or higher. 1536-well format is preferred. Assay volume 4-6 µL
Assay Steps	≤10 steps, including reagent additions, timed incubations, plate transfers to incubator, reading, etc.	≤10 steps including reagent additions, timed incubations, plate transfers to incubator, reading, etc.
Minimum time increments and maximum assay duration	Minimum assay window: 5 min. (i.e., earliest time point after last reagent addition)	< 24 h is ideal; max 48 h. Minimum assay window: 5 min.
Reagent Addition Steps	4 maximum (4 unique reagents max; more if pre-mixed)	4 maximum (4 unique reagents including cells max; more if pre-mixed)
*Reagent removal steps**	No plate coating steps	No aspiration steps
Temperature	Between RT and 37°C	Between RT and 37°C
*Demonstrated DMSO Tolerance**	0.5 – 1% DMSO	0.5-1% DMSO
Signal : Background Ratio	≥ 3-fold	≥ 3-fold
Day-to-Day variation of control (e.g., IC_{50}, EC_{50})	< 3-fold	< 3-fold
Reagent stability @ final working concentration	≥ 8 hrs @ RT or on ice bath; No on-line thawing	≥ 8 hrs @ RT or on ice bath; No on-line thawing
Validation run reagent supply	10 – 96-well plate equivalents	10 – 96-well plate equivalents
Protocol	Complete detailed protocol. All steps, equipment used, all vendor & catalog # for reagents. Data from 96-well or high density plate tests.	Complete detailed protocol. All steps, equipment used, all vendor & catalog # for reagents. Detailed cell culture procedure, passage #. Data from 96-well or high density plate tests.
Detectors	**PE ViewLux** (Top reading only: FI, TRF, FP, Abs, Luminescence) **MD Analyst** (bottom reading FI) **Acumen Explorer** (laser scanning imager)	**PE ViewLux** (Top reading only: FI, TRF, FP, Abs, Luminescence) **MD Analyst** (bottom reading FI) **Acumen Explorer** (laser scanning imager)
Special	For unique reagents, either investigator prepares sufficient quantity for HTS or identifies a reliable 3rd party vendor.	Cells must be certified micoplasma-free by direct culture assay and cell-DNA fluorochrome staining.

Abbreviations: # = number, Abs = absorbance, h = hour, FI = fluorescent intensity, FP = fluorescent polarization, RT = room temperature, TRF = time resolved fluorescence.

Plate Formats: 96-well plates contain 8 rows x 12 columns with volumes ranging between 50-200 µL; 384-well plates contain 16 rows x 24 columns with volumes ranging between 30-50 µL; 1536-well plates contain 32 rows by 48 columns with volumes ranging between 2-8 µL.

Reagent removal steps: Any step that requires the removal of material from the well of a microtiter plate. Such steps may be routine with 96-well plates; they are not recommended on robotic systems/1536-well plates.

Demonstrated DMSO Tolerance: Because all compounds screened are stored in ~100% DMSO and delivered as a 1 to 100 dilution to the assay, the sensitivity of the assay to between 0.5 and 1% DMSO must be determined.

One approach we used to gain information relevant to our screening strategy was to sponsor a Request for Information workshop on "High Throughput Screening Approaches for Toxicology" on September 11-12, 2008, at NIEHS. At this workshop, we sought input on existing commercial assays as well as assays under development, and technologies that might be exploited in future assay development efforts. Representatives from twenty-five companies gave presentations on technologies and assays that were potentially suited for use in toxicological profiling of large libraries of environmental compounds. There were 133 pre-registered attendees with a total attendance exceeding 150. Attendees were from large companies, small biotechnology firms, universities, advocacy organizations, and government agencies. Discussions provided us with a broad perspective on the types of assays that were currently available, the research trends for the future, and the opportunities for integrating varied screening approaches with the comprehensive screening program that NTP sought to develop. Assays or technologies presented at this workshop were considered by the Assay Working Group, and some have been implemented in our Tox21 efforts.

Another approach to acquire information pertaining to assay selection strategy, biological significance of the endpoints and pathways screened, and data interpretation was the convening in mid-2009 of a group of NIEHS/NTP scientists to serve as an advisory Faculty to the BSB and the Tox21 Assays Working Group (see list of members at the beginning of this section). The NTP HTS Faculty meets monthly. The main goals of the HTS faculty are to:

- implement the use of a variety of quantitative HTS approaches for identifying and characterizing toxic substances
- aid in identification of critical pathways and targets to investigate using an HTS approach
- help design focused, targeted studies using medium throughput approaches to collect data on defined groups of selected compounds (e.g., immunotoxicants, developmental toxicants, neurotoxicants).

In addition to broadly identifying assays and pathways, faculty meetings provide a forum for discussing the interpretation of data derived from qHTS as well as medium throughput and targeted screens. The Faculty also participates in discussions to develop a mechanism for formal reporting of HTS data, beyond deposition of the data in PubChem, NTP's CEBS, and U.S. EPA's ACToR.

III.3.2.1 Tox21 Phase I approach at the NCGC

The first qHTS screening activities at the NCGC focused on evaluation of a basic cytotoxicity response (measured as intracellular ATP levels) in 13 cell types (9 human, 2 rat, 2 mouse) (**Table III.3-2**). This screening was conducted to investigate responses among cell types and species, and to test the performance of a qHTS assay with our compound libraries screened over 14 concentrations (0.5 nM to 92 μM) using a 44-hour exposure duration. As anticipated, there were compounds that were cytotoxic or were not cytotoxic across all of the cell types at the concentrations tested. However, there were also compounds that were uniquely cytotoxic to one or more cell types and it was not possible to extrapolate results from one cell type or one species to another (Xia et al. 2007). These results appeared to be unrelated to cell turnover rates. In addition to considering species and tissue of origin, we also explored the use of primary cells versus established, commercially available cell lines. From a biological perspective, primary

Table III.3-2 Cell types investigated in assays conducted during Tox21 Phase I testing

Species	Cell type	Tissue of origin
Human	HEK 293	Kidney
	HepG2	Hepatoma
	SH-SY5Y	Neuroblastoma
	SK-N-SH	Neuroblastoma
	Jurkat	Acute T-cell leukemia
	BJ	Foreskin fibroblasts
	HUV-EC-C	Umbilical vein vascular endothelium
	MRC-5	Lung fibroblasts
	U2-OS	Osteosarcoma
	HeLa	Cervical carcinoma
	ME-180	Cervical carcinoma
	B lymphocytes	Lymphoblastoid cells
	HCT-116	Colon cancer
	Mesenchymal cell	Renal glomeruli
	THP-1	Monocytic leukemia
Chicken	B lymphocytes – multiple clones, each with a specific DNA repair deficiency	Blood
Hamster	CHO-K1	Chinese hamster ovary
Mouse	N2a	Neuroblastoma
	NIH3T3	Embryonic fibroblasts
Rat	H-4-II-E	Liver carcinoma
	Primary Renal Proximal tubule	Kidney, freshly isolated

cells might be preferred, but they present challenges in a qHTS 1536-well plate format because they are not as easily available, require special handling, and are not easily adaptable to 1536-well assay conditions.

These results impact our screening strategy for Phase II at the NCGC. Given that the response patterns determined for compounds tested across a range of qHTS assays at the NCGC will be used to prioritize compounds for more comprehensive *in vitro* and lower organism model systems (e.g., *Caenorhabditis elegans*, zebrafish embryos), one issue that has been extensively discussed is whether to restrict, to the extent possible, gene transactivation assays to the same cell type to reduce the number of variables impacting the results (and to potentially reduce the number of cytotoxicity assays that will need to be run to help explain the results obtained in gene transactivation assays) or to select each gene transactivation assay based solely on maximizing sensitivity and reproducibility. Both approaches have value; which approach will be used will largely depend on what assay versions are available (or could be constructed) and the extent to which an endpoint is cell-type specific.

Assay selection for Phase I at the NCGC has been constrained by the availability of suitable assays, both from a technological perspective and a biological perspective. The qHTS assays used in Phase I at the NCGC (**Appendix III.3-1**) were selected for technical conformation to NCGC requirements and for the generation of data considered relevant to toxicological effects. Essentially, Phase I screening at the NCGC was an ambitious pilot study established to evaluate assay performance (**Appendix III.3-2**) and cell type, methods of assay protocol optimization, and

the extent to which protocols could be varied without compromising the results. The qHTS data generated was also essential for the development of appropriate statistical analysis procedures to allow accurate, automated evaluation of thousands of qHTS concentration curves for the identification of actives and inactives. In addition, a number of strategies for following up initial screens to confirm and extend the information obtained were explored. During the last five years, the number of suitable assays has increased greatly and continues to expand, due at least in part to the interest that the Tox21 program has generated internationally among commercial enterprises and government/university research laboratories.

III.3.2.2 HTS Assays used during Tox21 Phase I testing

Assay selection was accomplished via several mechanisms. Initially, five commercial assays were used to test the responses seen with environmental chemicals in a qHTS robots platform. Previously, only chemicals of pharmaceutical interest (small molecules that met the Lipinski Rule of 5) had been tested in such a format. These initial assays were the Promega CellTiter Glo® cell viability assay; the Promega Caspase Glo® 3/7, 8, and 9 assays to measure induction of apoptosis; and the Promega CytoTox-ONE™ assay to measure cell death by detecting lactate dehydrogenase release.

Upon successful completion of these assays in multiple cell types, additional assays were added to the screening effort. These were variously nominated and supplied by scientists at the NIEHS/NTP, the U.S. EPA, and the NCGC, or they came through a formal public nomination process available via the NTP and EPA websites (http://ntp.niehs.nih.gov/go/27911, http://www.epa.gov/ncct/Tox21/, respectively) (**Figure III.3-1**).

In addition to Tox21 assays, the NTP library of 1408 compounds was occasionally included in screens conducted at the NCGC for the Molecular Libraries Screening Initiative (http://mli.nih.gov/mli/), an effort unrelated to Tox21 but one that provided potentially useful data. The list of these assays in which our NTP compounds were screened is provided in **Appendix III.3-4**. A comprehensive list of assays conducted at the NCGC in support of Tox21 and through the U.S. EPA ToxCast™ Phase I program where gene notations are possible is presented in **Appendix III.3-4**.

III.3.2.3 Results and lessons learned during Phase I of screening

During Phase I, a screening strategy was utilized that included assay optimization; the initial screen; data evaluation; identification of actives, negatives, and inconclusives; selection of a limited set of compounds for confirmational testing; and follow-up assays to confirm mode of action or target/pathway activity leading to identification of previously unknown toxicants or classes of compounds with a defined activity. The results of this approach have been published in journals such as Environmental Health Perspectives and Toxicological Sciences, and have been presented as poster and platform talks at national and international meetings. Additional publications are currently in preparation (see publication list). During these efforts, we learned the following lessons.

HTS Assay Nomination Form

Date:

Name of nominators:

Organization that nominator represents:

Contact Information for nominator (email and telephone):

Assay Name:

Biological/Toxicity Pathway:

Relevance of the pathway/target to Tox21:

Critical Factors for Assay Success:

Assay Technology:

Assay Source (and website if available):

Assay Format:

Reference Compounds:

Validation Status:

Major Estimated Costs:

Figure III.3-1 The Tox21 assay nomination form

1. Depending on the nature of the assay (e.g., stably transfected nuclear receptor construct versus stress response pathway reporter), each cell type might respond uniquely to a large set of compounds. Therefore, to aid in data interpretation and prioritization, a viability assay would ideally also be run for each assay that uses a different cell type or the same cell type but a different experimental condition. However, if an independent viability assay is run with each assay, primary assay throughput at the NCGC can be decreased by as much as 50% (the estimated assay throughput for Phase II at the NCGC is 25-37 assays per year). Three approaches would minimize any reduction is primary assay throughput:

    - The qHTS assays run at the NCGC are an initial screen designed to begin a triage process for the further testing of compounds identified as active in one or more assays. This decreases but does not eliminate the value of running independent cytotoxicity assays for assays that measure increasing signal as the response.

    - Assays that measure different endpoints in the same cell type under similar conditions (e.g., the same exposure duration) would reduce the number of independent cytotoxicity assays that would be needed.

- There is the potential, through commercial sources or by in-house developments, to use assays suitable for qHTS that contain multiple cell sensors, one that measures the activity of interest while the other serves to normalize for cell number and nonspecific effects (e.g., Davis et al. 2007).

2. Maximizing the acquisition of high quality, easily interpreted data (not to imply simplistic data) is essential to screening efficiency and is necessary for constructing a comprehensive profile of compound responses across a battery of different assays. Therefore, the results of optimization assays are carefully considered before moving an assay from the optimization stage to full production.

3. Inherent in any screening program is the trade-off between minimizing false negative or false positive rates; generally at some level of screening, it is not possible to minimize both. From a primary screening strategy point of view, our Phase II strategy for the NCGC is to select, to our best ability, assays that minimize the potential for false negatives. Given that, we appreciate fully, due to technical limitations inherent in the current qHTS paradigm (e.g., the absence of an exogenous metabolic activation system), that false negatives will occur. However, the compounds that are detected with high potency in one or more assays are likely to warrant further attention (i.e., they do not require metabolic transformation to produce an effect at physiologically relevant concentrations [i.e., <100 μM], against targets that are reflective of biological effects at the cellular level).

4. There are "critical" targets, such as the gap junctions involved in cell-to-cell communication or the orphan nuclear receptor CAR (constitutive androstane receptor), for which there is no existing *in vitro* assay or at least not one amenable to qHTS. Development of appropriate assays for these targets will be encouraged through contact with scientists in academia, government, and commercial enterprises, and for NTP through the release of SBIR grants and contracts (see Section III.3.3).

5. There is currently no method for including metabolic activation in the qHTS screens as the standard approach of using S9 mix from the liver of induced rats is toxic after a few hours and the homogeneous assays we use at NCGC cannot have any aspiration steps (see **Table III.3-1**). However, we have been evaluating other approaches for including xenobiotic metabolism in qHTS assays, including culturing primary hepatocytes alone or with a co-cultured reporter gene assay, culturing 3D liver model inserts into wells along with a co-cultured reporter gene assay, or using HepaRG cells as the target cell population.

The findings generated by Tox21 Phase I testing at the NCGC have demonstrated the applicability of the qHTS approach for toxicity screening of a large library of environmental compounds. Assays originally developed for drug discovery can be used, directly or with modification, to evaluate cellular processes potentially involved in toxicity responses. Statistical approaches have been developed to analyze the enormous amounts of data produced from qHTS screens. In some assays, data analysis has identified active compounds that have been further tested for confirmation of mechanism of action, and results are consistent with prior existing toxicity data from traditional *in vitro* or *in vivo* assays.

III.3.2.4 Tox21 Phase II NCGC qHTS screening strategy

To handle the 10K library tested in triplicate (effectively a 30K compound library), the NCGC has purchased a dedicated Kalypsys robotics facility to serve Tox21, with operation scheduled for early 2011. Prior to screening the Tox21 compound library in this facility, assays optimized for 1536-well format off-line will be evaluated for performance on the Kalypsys (i.e., on-line) using a smaller compound library, such as the Sigma-Aldrich 1280 compound LOPAC library (Library of Pharmacologically Active Compounds), run in triplicate. These evaluations will not impact on the overall screening schedule. If the qHTS assay performs as well on-line as off-line, it will be used to screen the Tox21 10K library. Criteria required for acceptable use include a Z factor greater 0.5, a CV less than 10%, and a signal to background ratio larger than 3. The methods of assay validation have been firmly established at NCGC using this procedure for assays used in the Molecular Libraries screening activities.

The capacity of the NCGC Tox21 laboratory is approximately 25 to 35 assays per year, depending on the complexity of the protocol and the assay duration. Taking NCGC assay throughput into account, the experience we have gained over the past 5 years, and the results of a comprehensive analysis of disease-associated cellular pathways (Golke et al. 2009), the current assay selection strategy focuses on a series of nuclear receptor assays and assays that measure the induction of stress response pathways in cells. A detailed description of the major stress response pathways and the rationale for selecting these as an important target for toxicological screening were recently published by Simmons et al. (2009). We believe that assays focused on activities involving these endpoints and pathways will indicate that a compound has entered a cell and induced a measurable response. Our strategy is designed to conduct as many assays as possible in a single cell type, when feasible, to aid in across-assay data comparisons for generation of a comprehensive toxicity profile for each compound and to potentially reduce the number of independent cell viability assays required to ascertain the cytotoxicity of our compounds in that assay. We anticipate that rapid, automated data analysis will allow effective monitoring of assay and compound library performance as we progress through our screening program and allow us to introduce modifications in assay selection and prioritization, protocol design, or cell type selection as needed. The current proposed strategy is one that we believe to be based on the best technology currently available and compatible with our qHTS 1536-well format. Given that, we view the assay panel and screening strategy as a flexible and changeable path forward that can be modified as new information, new assays, and new technologies become available.

In early July, 2010, we held two meetings attended by NIEHS and EPA intramural scientists, as well as academic colleagues who currently have collaborative Tox21 projects, to discuss qHTS assay screening strategies at the NCGC for nuclear receptor transactivation assays and assays that measure effects on stress response pathways. The purpose for those meetings was to get input from experts in the biology underlying our qHTS assays to ensure that we had not overlooked key factors in developing our assay strategy for screening the 10K library at the NCGC. The strategy outlined below is based on the discussions held at those meetings.

A set of nuclear receptor transactivation assays and stress response assays are proposed for initial Tox21 Phase II qHTS screening at the NCGC. Most of the specific endpoints and pathways were screened during Tox21 Phase I at the NCGC and where appropriate, the same assay will be used in Phase II. In some cases, although the endpoint/pathway is considered critically important for toxicological characterization, the assay that was used during Phase I did not perform as expected due, for example, to excessive variability in the signal, the signal to background ratio was not as large as expected, and/or the sensitivity of the assay was not as great as expected. Therefore, we will continue to search for better

performing assays for some of the endpoints/pathways we plan to screen and wherever a more technically acceptable assay can be identified and optimized, it will be used in Phase II.

In regards to cell type selection for constructs that can be put into any one of several cell types, we propose to use HepG2 cells as often as possible, based on both biological considerations and technical requirements of qHTS. First, they are an established, self-renewing cell line that reduces our reliance on animals/humans as a continuous source of primary cells. Admittedly, any established cell line provides this advantage over primary cells. Secondly, although not normal human cells, they are derived from human tissue and, most importantly, can be easily cultured in 1536-well format and are easily manipulated by transfection or viral transduction. Unlike most tumor/immortalized cell lines, HepG2 cells are p53 functional and are also wild-type (functional) for the other stress response pathways that we have investigated in Phase I screening. HepG2 cells are well-characterized and under the experimental conditions used for qHTS, these cells do not endogenously express high levels of the xenobiotic-metabolizing CYP450s. We are considering the feasibility of using HepaRG cells for that purpose.

III.3.2.5 Stress response assays

Our stress response assay screening strategy is based on the premise that compounds that induce one or more stress response pathways are more likely to exhibit *in vivo* toxicity than those that do not. Furthermore, by "binning" active compounds across the various stress response pathways, in relationship to active compounds with known *in vivo* toxicities, and by taking into account physical-chemical properties and various QSAR-based prediction models, we should be able to establish a prioritization scheme similar to the ToxPi one developed by EPA's NCCT for endocrine disruptors (Reif et al. 2010). The stress response pathways/assays in the proposed qHTS strategy include the following:

1. **Oxidative stress (ARE, Nrf2) (Activation of the Antioxidant Response Element via the Nrf2 signaling pathway).** Two different assays were run at NCGC during Phase I. One assay, obtained from Invitrogen uses a beta-lactamase readout, and the other, a luciferase readout; both are in HepG2 cells. The luciferase assay was created by Dr. Steve Simmons at the U.S. EPA using a lentivirus transfection process that allows creation of stable reporter cells in a variety of cell types. Because of the differences in the number of compounds detected by the two assays, we plan to use both HepG2-based assays to build in redundancy and to maximize detection of compounds capable of inducing oxidative stress.

2. **Hsp70 (Heat shock protein 70; HSE).** We have conducted this assay in two cell lines, a beta lactamase assay in HeLa cells obtained from Invitrogen and a luciferase assay in HepG2 cells created by Dr. Simmons at the U.S. EPA. We plan to use the luciferase assay in HepG2 cells because the technical characteristics of the two assays were similar and the luciferase assay allows for direct cross comparison of results with the ARE, Nrf2 assay in HepG2 cells.

3. **DNA damage response.** Three types of assays have been conducted at the NCGC during Phase I: a beta lactamase assay for p53 activation in HCT-116 cells obtained from Invitrogen, a luciferase assay for ELG1 DNA damage protein in HEK 293T cells developed by scientists at the NHGRI using a vector provided by Dr. Simmons at the EPA, and a differential cytotoxicity assay that compares cell viability (CellTiter Glo®) in several different DNA repair deficient DT40 chicken isogenic cell clones with that of the DNA repair competent parental (wild-type) cell line. These clones and the wild-type cell line were provided by Dr. Shunichi Takeda (Department of

Radiation Genetics, Kyoto University Medical School, Kyoto, Japan). Currently, comparative data analysis is underway for the ELG1 HEK 293T, p53 HCT-116, and chicken cell DNA damage assays. Once that analysis is complete, we can finalize assay selection for measuring DNA damage. Several factors will be considered in selecting these DNA damage assays. First, the ELG1 luciferase assay, currently in HEK 293T cells, can also be put into a HepG2 cell line (Simmons, EPA). HepG2 cells would provide uniformity across several assays but the relative sensitivities of the two assays would need to be compared and the HepG2 assay has not been optimized at NCGC. Second, although evaluating differential cytotoxicity across a broad battery of DT40 clones deficient in different DNA repair pathways is useful for potentially classifying the kind of DNA damage induced by a compound, this approach is resource intense and might not be needed for an initial screen. Therefore, we will select the two or three most sensitive cell lines (those that identify the greatest number of actives) and use them in the Phase II screening. Third, up-regulation of p53 occurs in response to a number of different cellular stresses, not just DNA damage, and therefore, actives identified in this assay are not necessarily DNA damaging agents, but may be inducing stress via other mechanisms.

4. **Mitochondrial damage assay (measurement of mitochondrial membrane potential) in HepG2 cells.** A fluorescent JC-10 water soluble dye is used to detect mitochondrial membrane depolarization, an indication of mitochondrial toxicity. The assay is fully optimized and technically acceptable. At the NCGC, this assay generated high quality data when used to screen the NTP 1408 compound library. The compounds detected as active in this assay were verified in a follow-up high content screen using the same dye. Studies on the active compounds are being conducted by the NTP using a Seahorse Biosciences SXF-24 Analyzer (http://www.seahorsebio.com/index.php) to evaluate how they affect mitochondrial function in HepG2 cells. We propose to include this assay in the Phase II screening program, as mitochondrial toxicity is implicated in a number of human disease conditions.

5. **AP-1 (Activator Protein 1) assay as a predictor of JNK, ERK, and p38 kinase signaling.** This signaling pathway responds to a variety of stimuli including inflammatory cytokines, oxidative stress, and infections. AP-1 also regulates important cellular functions such as cell proliferation, differentiation, and apoptosis. We propose to use an AP-1 luciferase stable reporter in HepG2 or HEK 293T cells for measuring AP-1 signaling. These cell lines were created using a lentiviral-based vector by Dr. Simmons.

6. **NFkB (inflammatory response; assay detects agonists and antagonists of the Toll-like receptor signaling pathway).** This assay measures a downstream inflammatory response that can be initiated through any of several pathways. We have this assay in three cell lines, a beta-lactamase assay in ME-180 cells from Invitrogen, a luciferase assay in HEK293 cells from Promega, and a luciferase assay in HepG2 cells from Dr. Simmons. We are in the process of analyzing the data generated on the NTP 1408 using these three cell lines prior to making a final decision on which version to use in the future.

7. **ESRE assay in HeLa cells (lipid damage, Endoplasmic Reticulum Stress Response element).** ER stress is associated with a number of pathologies including neurodegenerative disease, diabetes, and tumor growth under hypoxic conditions. The assay version that has been tested at NCGC is a beta-lactamase assay from Invitrogen and it is technically acceptable. Dr. Simmons (U.S. EPA) has developed a luciferase assay for this target but this assay has not yet been optimized in a 1536-well format. However, the EPA assay can be constructed in any preferred cell type, which may provide an advantage in reducing the number of cell viability assays required as well as

facilitate cross-assay comparison of results. There is a question of how much additional information an ESRE assay might provide, since it is expected that there will be significant overlap between the active compounds in this assay and actives identified in the Hsp70 assay. There may be some compounds that are uniquely active, however, and that might be of interest to us. Therefore, we gave this assay a lower priority and will consider whether to incorporate some version of it into the screening program after the higher priority assays have been completed and the data analyzed and interpreted.

8. **hERG potassium channel assay (human ether-a-go-go gene channel; plays a critical role in repolarization of cardiac action potential, and reduction in activity may lead to prolonged QT interval).** Screening compounds for hERG channel activity is an important method of determining potential cardiotoxicity. This assay from Invitrogen, in U-2 OS cells, must be run on a Hamamatsu machine, which will not be integrated into the Kalypsys robot. Therefore, the Tox21 library cannot be run in triplicate in this assay. However, it might be possible to run the library once, as a single set of the 10K compounds.

9. **Caspase 3/7 (irreversible induction of apoptosis).** There are opposing levels of enthusiasm for running this assay, with some Assay and Pathway Working Group members feeling that it would provide data to allow binning of compounds by mechanism of cell death and other members feeling that the assay is not sufficiently informative to justify resources. The assay can be run on any cell type and can be multiplexed with a cell viability assay, so there might be the possible of measuring both endpoints, depending on assay exposure duration.

10. **Cytotoxicity (measuring ATP levels using CellTiter Glo®).** Cell death represents the failure of cells to survive stress and represents a useful endpoint in and of itself for binning compounds. In addition, for assays that measure a decrease in signal (e.g., nuclear receptor antagonists assays), an assessment of cytotoxicity is critical for the accurate interpretation of data. Cytotoxicity assays, either using a second readout in the same well or an independent assay, will be conducted for all antagonist assays. However, we are also considering whether or not a cytotoxicity assay independent of any other readout would add value to the qHTS strategy for the 10K compound library. Issues under discussion include what cell type or types would provide the most useful data and under what assay conditions. This approach would be used only if we concluded that the cytotoxicity data collected to support other readouts would not be sufficient for classifying compounds according to their cytotoxic potency.

III.3.2.6 Nuclear receptor transactivation assays

Varying opinions have been expressed about the importance of screening full-length versus partial (ligand-binding domain [LBD]) nuclear receptors in transactivation assays, as there are advantages and disadvantages to each approach. On the one hand, there is concern that a truncated receptor will not only miss compounds that activate the nuclear receptor through an LBD-independent process but that the frequency of false positive will be increased due the greater access of some compounds to the LBD. On the other hand, concern was expressed that data generated with a full-length receptor would be more prone to false positives and/or difficulties in data interpretation. To resolve this issue, we decided to screen estrogen receptor alpha (ERα) in both formats and compare the results not only between the two assays but also with results obtained with other existing ERα transcriptional activation assays. Conclusions from this comparison may affect the strategies proposed for screening the other receptors. In addition, we plan, in the short-term, to limit nuclear receptor antagonist mode assays to ER and the androgen receptor (AR). Antagonist mode nuclear receptor assays can be confounded by the presence

of cytotoxicity and statistical and non-statistical approaches to distinguish between antagonism and cytotoxicity are under development. Conducting antagonist assays with other nuclear receptors will depend on our ability to correctly identify known ERα and AR antagonists using qHTS. The nuclear receptor assays listed below are prioritized in terms of sequence of qHTS screening at the NCGC; however, this is a draft scheduling priority and might change as screening results are obtained and interpreted.

1. **hERα (human estrogen receptor alpha).** The recommendation from our July 2010 Workshop on Nuclear Receptor Strategies was to use both a full-length and a partial (LBD) hERα assay in qHTS at the NCGC. Another recommendation was to use a receptor binding or co-activator recruitment assay. Two candidate full-length hERα assays are currently undergoing optimization studies at the NCGC; one uses a T47D-KBluc assay used by the U.S. EPA (Wilson et al. 2004), while the other uses BG-1 based assay. The BG1 assay was used in the NICEATM-sponsored validation of LumiCELL® (http://iccvam.niehs.nih.gov/methods/endocrine/endocrine.htm). The beta-lactamase reporter cell line with the GAL-4 system in HEK 293 cells from Invitrogen that was used during Phase I at the NCGC produced a relatively low signal to background ratio in 1536-well format and therefore, the data from the screen were difficult to interpret. Therefore, we are seeking an ERα LBD assay that provides a greater signal to noise ratio and thus better sensitivity. We are seeking also a hERβ assay. Discussions about a receptor binding or co-activator recruitment assay for qHTS continue.

2. **hAR (human androgen receptor).** As with the hER assay, we are currently evaluating additional assays using different technologies to make sure that we are using the most appropriate assay. The current partial hAR beta-lactamase assay with the GAL-4 system in HEK 293 cells from Invitrogen seemed to perform reasonably well when run in agonist mode (data are in the process of analysis) but considerable concerns were expressed at our July 2010 Workshop on Nuclear Receptor Strategies about the fact that the partial construct does not include the N-terminal end of the receptor, a region considered important for androgen responsiveness. We are currently attempting to optimize a cell line that contains a full length hAR, the MDA-kb-2 cell line used by the EPA (Wilson et al. 2001). Since this cell line also contains the glucocorticoid receptor (GR), we would need to conduct the assay with and without an AR antagonist to specifically identify AR agonism activity.

3. **hPPARγ (human peroxisome proliferator activated receptor gamma).** This assay (Invitrogen beta-lactamase reporter cell line with GAL-4 system in HEK 293 cells, with a partial human nuclear receptor construct) was used in Phase I at the NCGC. Technical performance was good and data quality was high. Therefore, we plan to use this assay and depending on the outcome of the comparative ER assays, we may continue to seek assays with a full-length receptor.

4. **hFXR (human farnesoid X receptor).** This is another Invitrogen beta-lactamase reporter cell line with GAL-4 system in HEK 293 cells, with a partial human nuclear receptor construct, that was successfully run in Phase I at the NCGC. Technical performance was good and data quality was high. Therefore, depending on the outcome of the comparative hERα assays, we plan to use this assay or identify an assay with a full-length receptor, if possible, in Phase II screening.

5. **hVDR (human vitamin D receptor).** This is another Invitrogen beta-lactamase reporter cell line with GAL-4 system in HEK 293 cells, with a partial human nuclear receptor construct, that was successfully run in Phase I at the NCGC. Technical performance was good and data quality was

high. Therefore, depending on the outcome of the comparative hERα assays, we plan to use this assay or identify an assay with a full-length receptor, if possible, in Phase II screening.

6. **hTRβ (human thyroid hormone receptor beta).** This is another Invitrogen beta-lactamase reporter cell line with GAL-4 system in HEK 293 cells, with a partial human nuclear receptor construct, that was successfully run in Phase I at the NCGC. Technical performance was good and data quality was high. We are also currently evaluating a luciferase reporter assay from Dr. A.J. Murk (Wageningen University, Netherlands) (Gutleb et al. 2005, Schriks et al. 2006). The assay consists of a stable reporter gene cell line developed using the thyroid hormone responsive rat pituitary tumor GH3 cells, that constitutively expresses both thyroid receptor isoforms (T3 and T4). An advantage of this full-length receptor assay is that Dr. Murk plans to have additional cell lines available with overexpressed CYP enzymes to provide the metabolic activation required for some known environmental chemical TR ligands. In addition, TRα and TRβ specific cell lines may be available soon from Dr. Murk's laboratory, which, if they can be optimized to a 1536-well format, will allow better extrapolation to *in vivo* effects. We plan to use one of these assays (the Invitrogen beta-lactamase reporter cell line with GAL-4 system or Dr. Murk's cells, depending on assay performance).

7. **hGR (human glucocorticoid receptor)**. The beta-lactamase assay we currently have from Invitrogen is in HeLa cells with a full-length human receptor under control of the mouse mammary tumor virus (MMTV) promoter. This cell line also contains an active AR and, therefore, additional runs will be needed to sort out which receptor is implicated in any observed responses. We have additional hGR cell lines provided by Dr. Trevor Archer (NIEHS) that await optimization and may provide an alternative screening approach for hGR active compounds.

8. **PXR (Pregnane X receptor).** We plan to test both rat and human PXR, as responses were markedly different between these two full-length receptors in Tox21 Phase I screening. This assay, from Puracyp, Inc. (Carlsbad, CA) is in HepG2 cells, with a luminescent readout and it gives acceptable responses.

9. **hAhR (human aryl hydrocarbon receptor).** This full-length receptor assay is in a HepG2 cell line with a luciferase reporter obtained from Puracyp, Inc. It was successfully run in Phase I screening and we plan to run this assay in Phase II.

10. **hLXR (human liver X receptor).** This partial human receptor beta-lactamase assay with a GAL-4 system in HEK 293 cells from Invitrogen was tested in Phase I screening; data quality did not meet required standards. Additional work to attempt to resolve the problems with this assay will be conducted. If the data quality problems cannot be resolved in additional experiments, we will evaluate other assays for this nuclear receptor.

11. **hRXR (human retinoid X receptor).** This is a valuable receptor to screen, but the HEK 293 cell line we currently have, which uses a partial receptor and a beta-lactamase readout, generated low quality data. We plan to investigate additional assays for this receptor.

12. **hPPARα (human peroxisome proliferator activated receptor alpha).** This partial-length receptor, luminescence read-out assay from DiscoveRx (Fremont, CA) has been optimized but not yet screened. However, due to the right-shifting of the agonist curve for the positive control

(GW7642), we will continue to search for more sensitive assays before making a final decision on whether to screen for this receptor and if so, which assay to use.

13. **hRAR (human retinoic acid receptor).** We have no previous experience at NCGC with an assay for this receptor. Prior to assay selection, we will consider the results of the hERα comparative study.

14. **hPR (human progesterone receptor).** We have no previous experience at NCGC with an assay for this receptor. Prior to assay selection, we will consider the results of the hERα comparative study.

15. **hPPARδ (human peroxisome proliferator activator receptor delta).** This nuclear receptor is not as biologically influential as PPARλ or α; we identified no actives in the Phase I run at the NCGC with the exception of compounds from the nuclear receptor reference ligand set. We therefore plan to conduct this assay only if resources and time permit.

III.3.2.7 ToxCast™ Phase II screens

In addition to the assays conducted at NCGC, 50 compounds from the NTP library will be included with the EPA compounds for screening in ToxCast™ Phase II. The companies that are participating in ToxCast™ Phase II, the assays they are planning to conduct, and the number of endpoints they will screen are presented in **Figure III.3-2**.

Figure III.3-2 ToxCast™ Phase II data sources

III.3.2.8 Additional screening activities initiated by the BSB during Tox21 Phase I

During Phase I, we arranged for the screening of small sets of compounds in selected assays to evaluate their utility (item 1), because the compounds being tested were not amenable to qHTS at the NCGC (item 2), or to further characterize a potentially interesting response obtained in qHTS at the NCGC (i.e., a follow-up study)(item 3).

1. HemoGenix, Inc. (Colorado Springs, CO). The HALO® platform for predictive stem cell hematoxicity screening was used to screen 25 compounds (known immunotoxicants, controls,

uncertain immunotoxicants) in six bone marrow progenitor cell types of human and mouse origin for differential cytotoxicity.

2. CertiChem, Inc. (Austin, TX). The MCF-7 ER Cell Proliferation Assay is being used to screen eight commercially available personal care products for estrogenicity and antiestrogenicity.

3. Integrated Oncology Solutions (Durham, NC). The Conformation Based Nuclear Receptor Ligand Profiling technology was used to screen six compounds for ER and AR agonist and antagonist activity in a hybrid cell line with a full length ER or AR, respectively.

III.3.3 Small Business Innovative Research (SBIR) Contracts

To accelerate the development of assays in areas of special interest to the NTP and Tox21, we released a set of SBIR contracts for award in fiscal year 2010 and then another set for award in fiscal year 2011. The contract topics included the following:

III.3.3.1 SBIR contracts for Fiscal Year 2010

The following assay-directed SBIR contract topics were made public in August of 2009 with an early November due date. After external review, contracts were awarded in early summer of 2010.

❖ **Development of Mid- to High-Throughput Toxicological Tests Using Model Organisms**

NTP is currently evaluating the utility of *C. elegans* as a model organism for toxicity testing and wishes to expand these capabilities by supporting the development of mid- to high- throughput alternative models that utilize lower vertebrates (e.g., fish) for evaluating the ability of substances of concern to the NTP to induce toxicological effects (e.g., developmental toxicity, reproductive toxicity, cardiotoxicity). This contract was awarded to Physical Sciences, Inc. (Andover, MA) to develop moderate throughput, non-invasive toxicological assays in zebrafish embryos.

❖ **Incorporation of Metabolism into Quantitative High Throughput Screening Assays**

One of the current limitations of the majority of *in vitro* HTS assays is the lack of hepatic or target organ metabolism in the testing strategy. Without an ability to integrate robust human and rodent hepatic or other target organ metabolism into the Tox21 HTS program, efforts to correlate *in vitro* test results with *in vivo* toxicities will continue to present challenges. To support these efforts, the goal of this SBIR is to develop high throughput assays (e.g., 384 or 1536-well) that incorporate human or rodent hepatic metabolic capability. Useful approaches might be based on (but are not limited to) (1) directly measuring gene expression or protein changes for critical pathways in primary hepatocytes, in stem cells that have been differentiated into hepatocyte-like cells, or in cell lines that retain appropriate kinds and levels of drug metabolizing activity; (2) transfecting such cells with reporter genes; or (3) co-cultivation of cells containing reporter genes with functional hepatocytes. This contract was awarded to High Throughput Genomics (Tucson, AZ).

❖ **Development of Quantitative High Throughput Screens for the Detection of Chemicals That Modulate Gap Junction Intercellular Communication**

The goal of this SBIR is to support the development of quantitative high throughput screens for the detection of chemicals that adversely alter gap junction activities. The ability of chemicals to modulate gap junctional intercellular communication is one area of special interest to the NTP. Gap junctions are clusters of intercellular channels connecting adjacent cells, which permit the direct exchange of ions and

small molecules between cells. A large number of connexin genes, many of which are tissue-specific, are involved in regulating gap junctional intercellular communication. Gap junctional intercellular coupling is required both for rapid signaling between electrically excitable cells and for the slower spread of intercellular second messenger signals between other cell types. Compound-induced changes in cell-to-cell communication via alteration in gap junctions may result from inappropriate alteration of connexin gene expression, a form of epigenetic toxicity. This contract was awarded to Detroit R&D (Detroit, MI).

III.3.3.2 SBIR contracts for Fiscal Year 2011

The following assay-directed SBIR contract topics were made public in August of 2010 with an early November due date. After external review, award of the contracts is expected in Spring, 2011.

❖ High Throughput Screening for Reactive Oxygen Species Mediating Toxicity

It is well known that the generation of reactive oxygen species (ROS) produced by chemical exposure can damage DNA, protein and lipids resulting in a variety of pathologies. Relevant species include hydrogen peroxide (H_2O_2), hydroxyl radicals (•OH), singlet oxygen ($^1O_2$), superoxide anion ($O_2^{-\cdot}$), hypochlorite anion (^-OCl), peroxyl radicals (ROO•) and others. Although superoxide dismutases, catalases, and peroxidases are usually efficient defenses against ROS, these defenses can be overwhelmed resulting in measurable ROS accumulation and toxicity. This SBIR is intended to support the development of quantitative high throughput or high content screening methods for the detection of various reactive oxygen species generated by some environmental toxicants. The methods may either generally detect ROS or selectively detect particular oxygen species. Linkage of ROS generation to specific subcellular organelles, to specific macromolecular effects such as protein or DNA damage, or other biological or toxicity endpoints is encouraged. Inclusion of positive controls for ROS assays that show assay detection limits and specificity are needed. These assays will be conducted at the NIH Chemical Genomics Center (NCGC) using a robotic platform that imposes specific requirements on the experimental design that can be employed in the quantitative high throughput screens conducted there.

❖ *In Vitro* 3D Tissue Models for Toxicity Testing

This SBIR is intended to foster the development of *in vitro* experimental systems capable of replicating major organ systems in humans, to be used for increased throughput and high data content screening of the mechanistic and toxicological effects of potential environmental toxicants. An emphasis is on developing systems that replicate key functions, such as the barrier function at initial sites of contact (skin, lungs, gastrointestinal tract) and metabolism/clearance (liver and kidneys), as well as the key target organs (respiratory, nervous, vasculature, and reproductive systems) that are most relevant to environmental health. These engineered tissues can be generated using biopsy, explanted, or excess transplant tissue or differentiated human stem cells and therefore the screening systems can be more relevant to human health than models based on experimental animal tissues. The 3D tissue model should be amenable to (1) 'omics technologies to identify biomarkers of exposure and response, including biomarkers at the pathway and network level, and (2) strategies for manipulating the genetic background of the culture system to study alterations in susceptibility to environmental factors resulting from genetic variation.

❖ Application of 'Omics Technologies to Rodent Formalin-Fixed, Paraffin Embedded Tissue Samples

The NTP Vision for the 21st Century is to move toxicology from a predominantly observational science at the level of disease-specific models to a predominantly predictive science focused on target-specific, mechanism-based, biological responses. Thus, the NTP is placing an increased emphasis on the use of alternative assays for targeting key pathways, molecular events, or processes linked to disease or injury, and has established a HTS program, representing a new paradigm in toxicological testing. One of the most effective ways of evaluating relationships between molecular pathways identified from studies using cultured cells exposed to environmental agents and disease is through the use of 'omics technologies on tissue samples obtained from *in vivo* toxicity studies. The NTP maintains one of the largest repositories in the world of formalin-fixed paraffin embedded (FFPE) tissue samples collected from nearly every GLP toxicity study carried out by the program over its 30-plus year history. Detailed pathology has been performed on all samples in the repository accompanied by serum chemistries and observational measures; however, very little is known about the molecular-level changes that parallel the pathology observed in these tissue samples. Recent technical developments allow for the successful extraction of DNA, RNA, or protein from FFPE samples for use in low dimensional molecular biology analyses. However, methods for the global assessment of changes related to these macromolecules are only starting to be developed. The purpose of this SBIR is to support the development of methods and tools that enable the use of FFPE tissues for Next-generation sequencing analysis of the genome, transcriptome, and epigenome. Effectiveness of developed methods will be determined by comparison to data generated using fresh frozen tissue.

III.3.4 Tox21-related Publications

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95(1):5-12.

Houck KA, Dix DJ, Judson RS, Kavlock RJ, Yang J, Berg EL. 2009. Profiling bioactivity of the ToxCast chemical library using BioMAP primary human cell systems. J Biomol Screen 14(9):1054-66. Epub 2009 Sep 22.

Houck KA, Kavlock RJ. 2008. Understanding mechanisms of toxicity: insights from drug discovery research. Toxicol Appl Pharmacol 227(2):163-78.

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118(4):485-92.

Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, et al. 2009. The toxicity data landscape for environmental chemicals. Environ Health Perspect 117(5):685-95.

Kavlock RJ, Austin CP, Tice RR. 2009. Toxicity testing in the 21st Century: Implications for human health risk assessment. Risk Anal 29(4):485-487.

Kavlock RJ, Reif DM, Richard AM, Rotroff DM, Romanov S, Medvedev A, et al. 2010. Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. Chem Res Toxicol 15;23(3):578-90.

Xia M, Huang R, Sun Y, Semenza GL, Aldred SF, Witt KL, et al. 2009. Identification of chemical compounds that induce 1 HIF-1α activity. Toxicol Sci 112(1): 153–163.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho M-H, et al. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect 116:284-291.

III.3.5 References

Davis RE, Zhang Y-Q, Southall N, Staudt LM, Austin CP, Inglese J, et al. 2007. A Cell-Based Assay for IκBα Stabilization Using a Two-Color Dual Luciferase-Based Sensor. Assay Drug Dev Technol 5(1):85-103.

Gohlke JM, Thomas R, Zhang Y, Rosenstein MC, Davis AP, Murphy C, et al. 2009. Genetic and environmental pathways to complex diseases. BMC Syst Biol. 3:46-61.

Gutleb AC, Meerts IA, Bergsma JH, Schriks M, Murk AJ. 2005. T-Screen as a tool to identify thyroid hormone receptor active compounds. Environ. Toxicol Pharmacol 19:231-38.

Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, et al. 2006. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci USA 103(31):11473-11478.

Judson RS, Martin MT, Reif DM, Houck KA, Knudsen TB, Rotroff DM, et al. 2010. Analysis of eight oil spill dispersants using rapid, in vitro tests for endocrine and other biological activity. Environ Sci Technol. 44(15):5979-5985.

Reif D, Martin M, Tan S, Houck K, Judson R, Richard A, et al. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ Health Perspect, published 08 Sep 2010, doi:10.1289/ehp.1002180.

Schriks M, Vrabie CM, Gutleb AC, Faassen EJ, Rietjens IM, Murk AJ. 2006. T-screen to quantify functional potentiating, antagonistic and thyroid hormone-like activities of poly halogenated aromatic hydrocarbons (PHAHs). Toxicol In Vitro 20**:**490-98.

Shukla SJ, Huang R, Austin CP, Xia M. In press. The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform. Drug Discov Today.

Simmons SO, Fan CY, Ramabhadran R. 2009. Cellular stress response pathway system as a sentinel ensemble in toxicological screening. Toxicol Sci 111(2):202-25.

Wilson VS, Bobseine K, Gray Jr LE. 2004. Development and characterization of a cell line that stably expresses an estrogen-responsive luciferase reporter for the detection of estrogen receptor agonist and antagonists. Toxicol Sci 81(1):69-77.

Wilson VS, Bobseine K, Lambright CR, Gray Jr LE. 2001. A Novel Cell Line, MDA-kb2, That Stably Expresses an Androgen- and Glucocorticoid-Responsive Reporter for the Detection of Hormone Receptor Agonists and Antagonists. Toxicol Sci 66(1):69-81.

Xia M, Huang R, Sun Y, Semenza GL, Aldred SF, Witt KL, et al. 2009. Identification of chemical compounds that induce HIF-1 alpha activity. Toxicol Sci 112(1):153-163.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect 116(3):284-291.

III.3.6 Appendices

This page left intentionally blank

Appendix III.3-1 qHTS Assays conducted during Tox21 Phase I at NCGC, organized alphabetically

TARGET	CELL LINES	SPECIES	TISSUE OF ORIGIN	ASSAY READOUT	TOXICITY PATHWAY	Libraries Tested
Androgen receptor, agonist mode	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Androgen receptor, antagonist mode	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
ATP content	Hek 293	Human	Embryonic kidney cells	Luminescence	Cytotoxicity	NTP, EPA
ATP content	HepG2	Human	Hepatocellular carcinoma	Luminescence	Cytotoxicity	NTP
ATP content	SH-SY5Y	Human	Neuroblastoma	Luminescence	Cytotoxicity	NTP
ATP content	Jurkat	Human	T-cell leukemia	Luminescence	Cytotoxicity	NTP
ATP content	BJ	Human	Foreskin fibroblasts	Luminescence	Cytotoxicity	NTP
ATP content	HUV-EC-C	Human	Vascular endothelial cells	Luminescence	Cytotoxicity	NTP
ATP content	SK-N-SH	Human	Neuroblastoma	Luminescence	Cytotoxicity	NTP
ATP content	MRC-5	Human	Lung fibroblasts	Luminescence	Cytotoxicity	NTP
ATP content	Mesenchymal cell	Human	Renal glomeruli	Luminescence	Cytotoxicity	NTP, EPA
ATP content	Proximal tubule cell	Rat	Freshly isolated from kidney	Luminescence	Cytotoxicity	NTP
ATP content	H-4-II-E	Rat	Hepatoma	Luminescence	Cytotoxicity	NTP
ATP content	N2a	Mouse	Neuroblastoma	Luminescence	Cytotoxicity	NTP
ATP content	NIH 3T3	Mouse	Fibroblasts, embryonic	Luminescence	Cytotoxicity	NTP
ATP content	P53-HCT-116	Human	Colon cancer	Luminescence	Cytotoxicity	NTP, EPA
ATP content	ARE-HepG2	Human	Hepatocellular carcinoma	Luminescence	Cytotoxicity	NTP, EPA
ATP content	NFkB-ME-180	Human	Cervical carcinoma	Luminescence	Cytotoxicity	NTP, EPA
ATP content	Human PXR-HepG2	Human	Hepatocellular carcinoma	Luminescence	Cytotoxicity	NTP, EPA
Caspase 3/7	Hek 293	Human	Embryonic kidney cells	Luminescence	Apoptosis	NTP
Caspase 3/7	HepG2	Human	Hepatocellular carcinoma	Luminescence	Apoptosis	NTP
Caspase 3/7	SH-SY5Y	Human	Neuroblastoma	Luminescence	Apoptosis	NTP
Caspase 3/7	Jurkat	Human	T-cell leukemia	Luminescence	Apoptosis	NTP
Caspase 3/7	BJ	Human	Foreskin fibroblasts	Luminescence	Apoptosis	NTP

TARGET	CELL LINES	SPECIES	TISSUE OF ORIGIN	ASSAY READOUT	TOXICITY PATHWAY	Libraries Tested
Caspase 3/7	HUV-EC-C	Human	Vascular endothelial cells	Luminescence	Apoptosis	NTP
Caspase 3/7	SK-N-SH	Human	Neuroblastoma	Luminescence	Apoptosis	NTP
Caspase 3/7	MRC-5	Human	Lung fibroblasts	Luminescence	Apoptosis	NTP
Caspase 3/7	Mesenchymal cell	Human	Renal glomeruli	Luminescence	Apoptosis	NTP
Caspase 3/7	Proximal tubule cell	Rat	Freshly isolated from kidney	Luminescence	Apoptosis	NTP
Caspase 3/7	H-4-II-E	Rat	Hepatoma	Luminescence	Apoptosis	NTP
Caspase 3/7	N2a	Mouse	Neuroblastoma	Luminescence	Apoptosis	NTP
Caspase 3/7	NIH 3T3	Mouse	Fibroblasts, embryonic	Luminescence	Apoptosis	NTP
Caspase 8	HepG2	Human	Hepatocellular carcinoma	Luminescence	Apoptosis	NTP, EPA
Caspase 8	Jurkat	Human	T-cell leukemia	Luminescence	Apoptosis	NTP, EPA
Caspase 8	NIH3T3	Mouse	Fibroblasts, embryonic	Luminescence	Apoptosis	NTP, EPA
Caspase 8	H4-II-E	Rat	Hepatoma	Luminescence	Apoptosis	NTP, EPA
Caspase 8	Hek 293	Human	Embryonic kidney cells	Luminescence	Apoptosis	NTP, EPA
Caspase 8	SHSY5Y	Human	Neuroblastoma	Luminescence	Apoptosis	NTP, EPA
Caspase 9	HepG2	Human	Hepatocellular carcinoma	Luminescence	Apoptosis	NTP, EPA
Caspase 9	Jurkat	Human	T-cell leukemia	Luminescence	Apoptosis	NTP, EPA
Caspase 9	NIH3T3	Mouse	Fibroblasts, embryonic	Luminescence	Apoptosis	NTP, EPA
Caspase 9	H4-II-E	Rat	Hepatoma	Luminescence	Apoptosis	NTP, EPA
Caspase 9	Hek 293	Human	Embryonic kidney cells	Luminescence	Apoptosis	NTP, EPA
Caspase 9	SHSY5Y	Human	Neuroblastoma	Luminescence	Apoptosis	NTP, EPA
CREB, agonist mode	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	CREB pathway	NTP
CREB, antagonist mode	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	CREB pathway	NTP
DNA repair (ATM(-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	NTP, EPA
DNA repair (ATM(-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	40 compds
DNA repair (FANCc (-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	NTP, EPA
DNA repair (FANCc (-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	40 compds
DNA repair (ku70/rad54 (-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	NTP, EPA

TARGET	CELL LINES	SPECIES	TISSUE OF ORIGIN	ASSAY READOUT	TOXICITY PATHWAY	Libraries Tested
DNA repair (ku70/rad54 (-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	40 compds
DNA repair (pol β (-/-) clone#1)	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	NTP, EPA
DNA repair (pol β (-/-) clone#1)	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	40 compds
DNA repair (pol β (-/-) clone#2)	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	NTP, EPA
DNA repair (pol β (-/-) clone#2)	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	40 compds
DNA repair (rev3 (-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	NTP, EPA
DNA repair (rev3 (-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	40 compds
DNA repair (ubc13 (-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	NTP, EPA
DNA repair (ubc13 (-/-))	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	40 compds
DNA repair (pol β (-/-))	MEF	Mouse	Embryonic fibroblast	Luminescence	DNA Repair	NTP
DNA repair (pol β (+/+))	MEF	Mouse	Embryonic fibroblast	Luminescence	DNA Repair	NTP
DNA repair (wild type)	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	NTP, EPA
DNA repair (wild type)	B lymphocyte	Chicken	B lymphocyte	Luminescence	DNA Repair	40 compds
DNA repair (XRCC(-/-))	MEF	Mouse	Embryonic fibroblast	Luminescence	DNA Repair	NTP
DNA repair (XRCC(+/+))	MEF	Mouse	Embryonic fibroblast	Luminescence	DNA Repair	NTP
ELG1	HEK293	Human	Embryonic kidney cells	Luciferase reporter	Stress signaling	NTP
ESRE	HeLa	Human	Cervical carcinoma	beta-lactamase reporter	ER stress pathway	NTP, EPA
Estrogen receptor alpha, agonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Estrogen receptor alpha, antagonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Farnesoid X receptor, agonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Farnesoid X receptor, antagonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Fluorescent spectra scan for EPA and NTP sets						NTP, EPA
Glucocorticoid receptor, agonist	HeLa	Human	Cervical carcinoma	beta-lactamase reporter	NR signaling	NTP, EPA
Glucocorticoid receptor, antagonist	HeLa	Human	Cervical carcinoma	beta-lactamase reporter	NR signaling	NTP, EPA

TARGET	CELL LINES	SPECIES	TISSUE OF ORIGIN	ASSAY READOUT	TOXICITY PATHWAY	Libraries Tested
HapMap, genetics	74 cell lines	Human	Lymphoblast	Luminescence	Cytotoxicity	240 compds
Heat shock response	HeLa	Human	Cervical carcinoma	beta-lactamase reporter	Hsp signaling	NTP, EPA
hERG	U2OS	Human	Osteosarcoma	Thallium influx assay	K channel	NTP
Hsp70	HepG2	Human	Hepatocellular carcinoma	Luciferase reporter	Hsp70 signaling	NTP, EPA
Human AhR	HepG2	Human	Hepatocellular carcinoma	Luciferase reporter	NR signaling	NTP, EPA
Human PXR	HepG2	Human	Hepatocellular carcinoma	Luciferase reporter	NR signaling	NTP, EPA
Hypoxia, agonist	ME-180	Human	Cervical carcinoma	beta-lactamase reporter	HRE signaling	NTP, EPA
IL8	THP-1	Human	monocyte	HTRF (TR-FRET), Cisbio	Cytokines	NTP
Liver X receptor beta	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	
Liver X receptor beta, agonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Liver X receptor beta, antagonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
LXRbeta protein interaction	CHO K1	Hamster	Chinese hamster ovary	Enzyme fragment complementation	NR signaling	
Membrane integrity	Hek 293	Human	Embryonic kidney cells	LDH release	Cytotoxicity	NTP
Membrane integrity	Mesenchymal cell	Human	Renal glomeruli	LDH release	Cytotoxicity	NTP
Membrane integrity	Hek 293	Human	Embryonic kidney cells	Protease release	Cytotoxicity	NTP
Membrane integrity	Mesenchymal cell	Human	Renal glomeruli	Protease release	Cytotoxicity	NTP
Mitochondrial membrane potential	HepG2	Human	Hepatocellular carcinoma	Fluorescence	Cytotoxicity	NTP
NFkB, agonist	ME-180	Human	Cervical carcinoma	beta-lactamase reporter	NFkB signaling	NTP, EPA
NFkB, antagonist	ME-180	Human	Cervical carcinoma	beta-lactamase reporter	NFkB signaling	NTP, EPA
Nrf2	HepG2	Human	Hepatocellular carcinoma	Luciferase reporter	Nrf2ARE signaling	NTP, EPA
Nrf2/ARE	HepG2	Human	Hepatocellular carcinoma	beta-lactamase reporter	Nrf2/ARE signaling	NTP, EPA
P53	HCT-116	Human	Colon cancer	beta-lactamase reporter	p53 signaling	NTP, EPA
Peroxisome Proliferator-activated receptor-delta, agonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA

TARGET	CELL LINES	SPECIES	TISSUE OF ORIGIN	ASSAY READOUT	TOXICITY PATHWAY	Libraries Tested
Peroxisome Proliferator-activated receptor-delta, antagonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Peroxisome Proliferator-activated receptor-gamma, agonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Peroxisome Proliferator-activated receptor-gamma, antagonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
PPARa protein interaction	CHO K1	Hamster	Chinese hamster ovary	Luminescence	Cytotoxicity	NTP, EPA
PPARa protein interaction agonist mode	CHO K1	Hamster	Chinese hamster ovary	Enzyme fragment complementation	NR signaling	EPA
PPARa protein interaction antagonist mode	CHO K1	Hamster	Chinese hamster ovary	Enzyme fragment complementation	NR signaling	EPA
Rat PXR	HepG2	Human	Hepatocellular carcinoma	Luciferase reporter	NR signaling	NTP, EPA
Retinoid X receptor alpha, agonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Retinoid X receptor alpha, antagonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
RORr	CHO	Hamster	Chinese hamster ovary	luciferase reporter	NR signaling	NTP
Thyroid hormone receptor beta, agonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Thyroid hormone receptor beta, antagonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
TNFa	THP-1	Human	monocyte	HTRF (TR-FRET), Cisbio	Cytokines	NTP
Vitamin D receptor, agonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA
Vitamin D receptor, antagonist	Hek 293	Human	Embryonic kidney cells	beta-lactamase reporter	NR signaling	NTP, EPA

Appendix III.3-2A 1536-well plate statistics and positive controls for each nuclear receptor assay

Assay	S/B	Z'	CV	Positive Control (μM)
AR agonist	2.2±0.2	0.60±0.10	10±2	R1881 (0.03)
ER agonist	2.5±0.5	0.64±0.14	11±4	17β-estradiol (0.02)
FXR agonist	5.4±0.5	0.68±0.08	9±4	Chenodeoxycholic acid (60)
GR agonist	4.2±0.5	0.59±0.11	16±2	Dexamethasone (0.1)
LXR agonist	3.0±0.2	0.76±0.11	9±4	T0901317 (15)
PPARδ agonist	2.4±2.5	0.75±0.04	9±3	L165041 (10)
PPARγ agonist	3.4±0.2	0.89±0.02	7±2	Rosiglitazone (2)
RXR agonist	1.9±0.1	0.74±0.05	8±1	Retinoic acid (0.5)
TRβ agonist	2.8±0.5	0.51±0.19	16±2	T3 (0.01)
VDR agonist	4.3±0.5	0.83±0.05	7±1	1α,25-Dihydroxy vitaminD3 (0.05)
AR antagonist	2.5±0.8	0.65±0.07	15±1	R1881 (0.01)
ER antagonist	2.7±1.1	0.43±0.08	9±0.4	17β-estradiol (0.0005)
FXR antagonist	4.9±0.4	0.73±0.05	7±1	Chenodeoxycholic acid (50)
GR antagonist	2.8±0.2	0.49±0.07	10±1	Dexamethasone (0.002)
LXR antagonist	3.5±0.6	0.62±0.38	12±3	T0901317 (1.5)
PPARδ antagonist	1.8±0.1	0.49±0.06	6±1	L165041 (0.5)
PPARγ antagonist	2.7±0.2	0.84±0.07	3±0.3	Rosiglitazone (0.05)
RXR antagonist	2.1±0.2	0.78±0.07	8±3	Retinoic acid (0.1)
TRβ antagonist	3.0±0.5	0.55±0.11	7±1	T3 (0.0004)
VDR antagonist	4.2±0.3	0.87±0.03	4±1	1α,25-Dihydroxy vitaminD3 (0.003)

Abbreviations: CV = coefficient of variation; S/B = signal to background; T3 = triiodothyronine
Z'-factor: the calculated separation band (distance between 3 standard deviations from the mean of the positive control signal and 3 standard deviations from the mean of the negative control signal) divided by the calculated assay dynamic range of the assay (distance between mean of the positive control signal and mean of the control measurements). The Z'-factor is the same as the Z-factor, using the positive control instead of the assay signal intensity.

Appendix III.3-2B **1536-well plate statistics and positive controls for each stress pathway assay**

Assay	S/B	Z'	Positive Control (µM)*
ARE bla, agonist	3.4±0.4	0.62±0.07	Beta-Naphthoflavone (23 and 46)
HSE luc, agonist	4.8±1.4	0.35±0.12	17-AAG (5.75 and 2.88)
HSE bla, agonist	3.5±0.4	0.72±0.06	17-AAG (0.61 and 0.5)
ER stress bla, agonist	2.8±0.1	0.81±0.03	17-AAG (3.83 and 1.92)
ARE luc	15±01.3	0.67±0.06	Beta-Naphthoflavone (5.8 and 11.5)
NFkB bla, agonist	6.9±21.6	0.46±0.33	TNFα (10 and 20 ng/ml)
NFkB bla, antagonist	8.2±1.7	0.87±0.04	TNFα (1 ng/ml)
P53 bla, agonist	3.9±0.2	0.64±0.04	Nutlin-3 (11.5 and 23)
HRE bla, agonist	6.4±1.1	0.71±0.11	CoCl2 (50 and 100)

Abbreviations: 17-AAG = 17-Allylamino-17-demethoxygeldanamycin; S/B = signal to background
Z'-factor: the calculated separation band (distance between 3 standard deviations from the mean of the positive control signal and 3 standard deviations from the mean of the negative control signal) divided by the calculated assay dynamic range of the assay (distance between mean of the positive control signal and mean of the control measurements). The Z'-factor is the same as the Z-factor, using the positive control instead of the assay signal intensity.
*There are two concentrations for positive control for the purpose of data normalization.

Appendix III.3-2C 1536-well plate statistics and positive controls for ELG1, mitochondrial, caspase 3/7, and hERG assays

Assay	S/B	Z'	Positive Control (μM)*
ELG1	5.2±0.5	0.73±0.15	MMS (600 and 700)
Mitochondrial, 1 hr	22.7±2.4	0.44±0.07	FCCP (3.5, 6.9 and 9.2)
Mitochondrial, 5 hr	29.7±2.8	0.77±0.12	FCCP (3.5, 6.9 and 9.2)
hERG	4.6±0.2	0.77±0.04	Astemizole (5 and 10)
Jurkat, Caspase 3/7	7.2±0.5	0.86±0.03	Tamoxifen (100)
Hek 293, Caspase 3/7	8.7±21.0	0.46±0.15	Tamoxifen (100)
BJ, Caspase 3/7	9.4±02.1	0.87±0.02	Staurosporine (10)
MRC5, Caspase 3/7	9.4±0.7	0.82±0.05	Staurosporine (10)
SK-N-SH, Caspase 3/7	11.9±0.5	0.87±0.02	Staurosporine (10)
SH-SY5Y, Caspase 3/7	7.9±0.8	0.80±0.04	Staurosporine (10)
N2a, Caspase 3/7	5.5±0.2	0.86±0.05	Tamoxifen (100)
3T3, Caspase 3/7	6.7±2.7	0.62±0.34	Tamoxifen (100)
H4-II-E, Caspase 3/7	14.1±0.6	0.46±0.26	Tamoxifen (100)
Mensangial, Caspase 3/7	16.4±0.6	0.84±0.03	Tamoxifen (100)
HUVEC, Caspase 3/7	8.6±0.7	0.52±0.11	Tamoxifen (100)
Kidney proximal tubule, Caspase 3/7	2.0±0.1	-0.006±0.14	Tamoxifen (100)
HepG2, Caspase 3/7	16.9±1.1	0.81±0.05	Tamoxifen (100)

Abbreviations: FCCP = carbonyl cyanide *p*-(trifluoromethoxy)phenylhydrazone]; MMS = methyl methane sulfonate; S/B = signal to background

Z'-factor: the calculated separation band (distance between 3 standard deviations from the mean of the positive control signal and 3 standard deviations from the mean of the negative control signal) divided by the calculated assay dynamic range of the assay (distance between mean of the positive control signal and mean of the control measurements). The Z'-factor is the same as the Z-factor, using the positive control instead of the assay signal intensity.

*There are two concentrations for positive control for the purpose of data normalization.

Appendix III.3-3 Additional assays conducted at the NCGC with the NTP and/or U.S. EPA libraries through the Molecular Libraries Initiative screening effort

Target (s) / Biology	Assay Category	PubChem AID	# Compds Screened
12-Lipoxygenase (12hLO)	Isolated Molecular Target	1452	
15-Hydroxyprostaglandin dehydrogenase	Isolated Molecular Target	894	
15-Lipoxygenase 1 (15hLO1)	Isolated Molecular Target	pending	78,081
15-Lipoxygenase 2 (15hLO2)	Isolated Molecular Target	881	
Aldehyde dehydrogenase A1 (ALDHA1)	Isolated Molecular Target	1030	
Anthrax intoxication pathway	Cellular Signaling, Sensor	912	40,513
APE1	Isolated Molecular Target	1705	
beta-glucocerebrosidase S1, S2	Isolated Molecular Target	360	59,823
beta-lactamase	Isolated Molecular Target	584	71,982
BRCT-pSXXF (GREEN)	Isolated Molecular Target	pending	78,845
BRCT-pSXXF (RED)	Isolated Molecular Target	pending	78,845
Caspase-1	Isolated Molecular Target	900	80,124
Caspase-7	Isolated Molecular Target	889	80,037
CBFβ-RUNX1 interaction blockers	Isolated Molecular Target	1484	
Cell based assay for hERG channel blockers	Cellular Signaling, Sensor	pending	
Cell signaling AP-1 BLA	Cellular Signaling, Reporter gene	357	76,644
Cell signaling HRE BLA	Cellular Signaling, Reporter gene	915	81,956
Cellular Toxicity, TTP antagonists	Toxicity	pending	
Chromosomal Remodeling (LDR)	Cellular Signaling, Reporter gene	597	69,135
Cruzain, Trypanosoma cruzi	Isolated Molecular Target	1476	
Endotoxin (TTP) agonists/antagonists	Cellular Signaling, Sensor	pending	
ERK Phosphorlyation	Cellular Pathway, ALPHA	995	
G alpha i-RGS	Isolated Molecular Target	880	
Hsp90 co-chaperone interaction	Isolated Molecular Target	595	73,422
Human alpha-galactosidase	Isolated Molecular Target	1467	
Human RECQ1 helicase	Isolated Molecular Target	2549	
Human tyrosyl-DNA phosphodiesterase I	Isolated Molecular Target	pending	
Huntington polyglutamine expansion-GFP/ATP	Cellular viablity	1471	56,494
Imprinting	Cellular Signaling, Reporter gene	pending	

Target (s) / Biology	Assay Category	PubChem AID	# Compds Screened
Inositol Monophosphatase Inhibitors	Isolated Molecular Target	901	
JNK3 activation	Cellular Pathway, ALPHA	530	11,210
mRNA Splicing thalassemia	Cellular Signaling, Splicing reporter	925	71,683
Multi-protein DNA Replication System	Isolated Molecular Target	603	73,892
Neuropeptide S receptor antagonists	Cellular Signaling, Sensor	1464	
O-Glc NAc Transferase	Isolated Molecular Target	447	70,308
Oxidoreductase HADH2	Isolated Molecular Target	886	85,316
Oxidoreductase HSD17b4	Isolated Molecular Target	893	78,845
P450 (CYP 1A2, Luc)	Profiling, Chemical Library	410	47,239
P450 (CYP 2C19, Luc)	Profiling, Chemical Library	899	
P450 (CYP 2C9, Luc)	Profiling, Chemical Library	883	11,703
P450 (CYP 2D6, Luc)	Profiling, Chemical Library	891	6,832
P450 (CYP 3A4, Luc)	Profiling, Chemical Library	884	19,727
p53 two temp., synthetic lethal	Cellular viablity	902	124,570
Potentiators of CRE signaling	Cellular Signaling, Reporter gene	662	93,601
Profiling for detergent-sensitive inhibitors	Profiling, Chemical Library	585	71,982
PXR binding	Isolated Molecular Target	pending	
Pyruvate Kinase, Leishmania	Isolated Molecular Target	1721	
Rice Alpha Glucosidase	Isolated Molecular Target	1466	
RORγ antagonists	Cellular Signaling, Reporter gene	2551	
Schistosoma Peroxiredoxins	Isolated Molecular Target	448	70,308
SF1 Receptor Agonists	Cellular Signaling, Reporter gene	pending	
SMN2 Splicing Modulators	Cellular Signaling, Reporter gene	1458	
Tau polymerization	Isolated Molecular Target	596	71,982
TRβ antagonists	Isolated Molecular Target	1469	
TSH Receptor	Cellular Signaling, Sensor	926	84,122
Ubiquitin Pathway, Ubiquitin-GFP	Cellular Pathway, Phenotypic	526	71,977
VP-16 cell counter screen for RORγ antagonists	Cellular Signaling, Reporter gene	2546	
YjeE:ADP binding	Isolated Molecular Target	605	71,977

Appendix III.3-4 Comprehensive list of Tox21 assays (NCGC and ToxCast™), where gene annotations are possible.

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	ATG_Ahr_CIS	AHR	aryl hydrocarbon receptor
EPA	TOXCAST	ATG_AP_1_CIS	FOSB	FBJ murine osteosarcoma viral oncogene homolog B
EPA	TOXCAST	ATG_AP_1_CIS	JUND	jun D proto-oncogene
EPA	TOXCAST	ATG_AP_1_CIS	JUN	jun oncogene
EPA	TOXCAST	ATG_AP_1_CIS	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
EPA	TOXCAST	ATG_AP_2_CIS	TFAP2A	transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)
EPA	TOXCAST	ATG_AP_2_CIS	TFAP2B	transcription factor AP-2 beta (activating enhancer binding protein 2 beta)
EPA	TOXCAST	ATG_AP_2_CIS	TFAP2D	transcription factor AP-2 delta (activating enhancer binding protein 2 delta)
EPA	TOXCAST	ATG_AP_2_CIS	TFAP2E	transcription factor AP-2 epsilon (activating enhancer binding protein 2 epsilon)
EPA	TOXCAST	ATG_AP_2_CIS	TFAP2C	transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)
EPA	TOXCAST	ATG_AR_TRANS	AR	androgen receptor
EPA	TOXCAST	ATG_BRE_CIS	SMAD1	SMAD family member 1
EPA	TOXCAST	ATG_BRE_CIS	SMAD4	SMAD family member 4
EPA	TOXCAST	ATG_BRE_CIS	SMAD5	SMAD family member 5
EPA	TOXCAST	ATG_BRE_CIS	SMAD6	SMAD family member 6
EPA	TOXCAST	ATG_C_EBP_CIS	CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha
EPA	TOXCAST	ATG_C_EBP_CIS	CEBPB	CCAAT/enhancer binding protein (C/EBP), beta
EPA	TOXCAST	ATG_C_EBP_CIS	CEBPD	CCAAT/enhancer binding protein (C/EBP), delta
EPA	TOXCAST	ATG_C_EBP_CIS	CEBPE	CCAAT/enhancer binding protein (C/EBP), epsilon
EPA	TOXCAST	ATG_C_EBP_CIS	CEBPG	CCAAT/enhancer binding protein (C/EBP), gamma
EPA	TOXCAST	ATG_CAR_TRANS	NR1I3	nuclear receptor subfamily 1, group I, member 3 (Constitutive Androstane Receptor)
EPA	TOXCAST	ATG_CMV_CIS	ATF2	activating transcription factor 2
EPA	TOXCAST	ATG_CMV_CIS	ATF4	activating transcription factor 4 (tax-responsive enhancer element B67); activating transcription factor 4C
EPA	TOXCAST	ATG_CMV_CIS	CREB1	cAMP responsive element binding protein 1
EPA	TOXCAST	ATG_CMV_CIS	CREB3	cAMP responsive element binding protein 3
EPA	TOXCAST	ATG_CMV_CIS	FOSB	FBJ murine osteosarcoma viral oncogene homolog B
EPA	TOXCAST	ATG_CMV_CIS	JUNB	jun B proto-oncogene
EPA	TOXCAST	ATG_CMV_CIS	JUND	jun D proto-oncogene

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	ATG_CMV_CIS	JUN	jun oncogene
EPA	TOXCAST	ATG_CMV_CIS	NFIA	nuclear factor I/A
EPA	TOXCAST	ATG_CMV_CIS	NFIB	nuclear factor I/B
EPA	TOXCAST	ATG_CMV_CIS	NFIC	nuclear factor I/C
EPA	TOXCAST	ATG_CMV_CIS	NFIX	nuclear factor I/X
EPA	TOXCAST	ATG_CMV_CIS	NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
EPA	TOXCAST	ATG_CMV_CIS	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
EPA	TOXCAST	ATG_CMV_CIS	RELA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)
EPA	TOXCAST	ATG_CMV_CIS	YY1	YY1 transcription factor
EPA	TOXCAST	ATG_CRE_CIS	ATF2	activating transcription factor 2
EPA	TOXCAST	ATG_CRE_CIS	ATF4	activating transcription factor 4 (tax-responsive enhancer element B67); activating transcription factor 4C
EPA	TOXCAST	ATG_CRE_CIS	CREB3	cAMP responsive element binding protein 3
EPA	TOXCAST	ATG_CRE_CIS	CREB5	cAMP responsive element binding protein 5
EPA	TOXCAST	ATG_CRE_CIS	CREM	cAMP responsive element modulator
EPA	TOXCAST	ATG_DR4_LXR_CIS	NR1H2	nuclear receptor subfamily 1, group H, member 2(Liver X receptor beta)
EPA	TOXCAST	ATG_DR4_LXR_CIS	NR1H3	nuclear receptor subfamily 1, group H, member 3(Liver X receptor alpha)
EPA	TOXCAST	ATG_DR5_CIS	RARA	retinoic acid receptor, alpha
EPA	TOXCAST	ATG_DR5_CIS	RARB	retinoic acid receptor, beta
EPA	TOXCAST	ATG_DR5_CIS	RARG	retinoic acid receptor, gamma
EPA	TOXCAST	ATG_E_Box_CIS	MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
EPA	TOXCAST	ATG_E2F_CIS	E2F1	E2F transcription factor 1
EPA	TOXCAST	ATG_E2F_CIS	E2F2	E2F transcription factor 2
EPA	TOXCAST	ATG_E2F_CIS	E2F3	E2F transcription factor 3
EPA	TOXCAST	ATG_E2F_CIS	E2F4	E2F transcription factor 4, p107/p130-binding
EPA	TOXCAST	ATG_E2F_CIS	E2F5	E2F transcription factor 5, p130-binding
EPA	TOXCAST	ATG_E2F_CIS	E2F6	E2F transcription factor 6
EPA	TOXCAST	ATG_E2F_CIS	E2F7	E2F transcription factor 7
EPA	TOXCAST	ATG_EGR_CIS	EGR1	early growth response 1
EPA	TOXCAST	ATG_EGR_CIS	EGR2	early growth response 2
EPA	TOXCAST	ATG_EGR_CIS	EGR3	early growth response 3
EPA	TOXCAST	ATG_ERa_TRANS	ESR1	estrogen receptor 1
EPA	TOXCAST	ATG_ERE_CIS	ESR1	estrogen receptor 1

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	ATG_ERRa_TRANS	ESRRA	estrogen-related receptor alpha
EPA	TOXCAST	ATG_ERRg_TRANS	ESRRG	estrogen-related receptor gamma
EPA	TOXCAST	ATG_Ets_CIS	ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)
EPA	TOXCAST	ATG_Ets_CIS	ETS2	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)
EPA	TOXCAST	ATG_FoxA2_CIS	FoxA2	forkhead box A2
EPA	TOXCAST	ATG_FoxO_CIS	FoxO1	forkhead box O1
EPA	TOXCAST	ATG_FoxO_CIS	FoxO3	forkhead box O3; forkhead box O3B pseudogene
EPA	TOXCAST	ATG_FXR_TRANS	NR1H4	nuclear receptor subfamily 1, group H, member 4 (Farnesoid X Receptor)
EPA	TOXCAST	ATG_GATA_CIS	GATA1	GATA binding protein 1 (globin transcription factor 1)
EPA	TOXCAST	ATG_GATA_CIS	GATA2	GATA binding protein 2
EPA	TOXCAST	ATG_GATA_CIS	GATA3	GATA binding protein 3
EPA	TOXCAST	ATG_GATA_CIS	GATA4	GATA binding protein 4
EPA	TOXCAST	ATG_GATA_CIS	GATA5	GATA binding protein 5
EPA	TOXCAST	ATG_GATA_CIS	GATA6	GATA binding protein 6
EPA	TOXCAST	ATG_GLI_CIS	GLI1	GLI family zinc finger 1
EPA	TOXCAST	ATG_GLI_CIS	GLI2	GLI family zinc finger 2
EPA	TOXCAST	ATG_GLI_CIS	GLI3	GLI family zinc finger 3
EPA	TOXCAST	ATG_GLI_CIS	GLI4	GLI family zinc finger 4
EPA	TOXCAST	ATG_GR_TRANS	NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
EPA	TOXCAST	ATG_GRE_CIS	NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
EPA	TOXCAST	ATG_HIF1a_CIS	HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)
EPA	TOXCAST	ATG_HNF4a_TRANS	HNF4A	hepatocyte nuclear factor 4, alpha
EPA	TOXCAST	ATG_HNF6_CIS	ONECUT1	one cut domain, family member 1
EPA	TOXCAST	ATG_HSE_CIS	HSF1	heat shock transcription factor 1
EPA	TOXCAST	ATG_HSE_CIS	HSF2	heat shock transcription factor 2
EPA	TOXCAST	ATG_HSE_CIS	HSF3	heat shock transcription factor 3
EPA	TOXCAST	ATG_HSE_CIS	HSF4	heat shock transcription factor 4
EPA	TOXCAST	ATG_HSE_CIS	HSF5	heat shock transcription factor family member 5
EPA	TOXCAST	ATG_IR1_CIS	NR1H4	nuclear receptor subfamily 1, group H, member 4 (Farnesoid X Receptor)
EPA	TOXCAST	ATG_ISRE_CIS	IRF1	interferon regulatory factor 1
EPA	TOXCAST	ATG_ISRE_CIS	IRF2	interferon regulatory factor 2

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	ATG_ISRE_CIS	IRF3	interferon regulatory factor 3
EPA	TOXCAST	ATG_LXRa_TRANS	NR1H3	nuclear receptor subfamily 1, group H, member 3(Liver X receptor alpha)
EPA	TOXCAST	ATG_LXRb_TRANS	NR1H2	nuclear receptor subfamily 1, group H, member 2(Liver X receptor beta)
EPA	TOXCAST	ATG_MRE_CIS	MTF2	metal response element binding transcription factor 2
EPA	TOXCAST	ATG_MRE_CIS	MTF1	metal-regulatory transcription factor 1
EPA	TOXCAST	ATG_Myb_CIS	MYB	v-myb myeloblastosis viral oncogene homolog (avian)
EPA	TOXCAST	ATG_Myc_CIS	MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
EPA	TOXCAST	ATG_NF_kB_CIS	NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
EPA	TOXCAST	ATG_NFI_CIS	NFIA	nuclear factor I/A
EPA	TOXCAST	ATG_NFI_CIS	NFIB	nuclear factor I/B
EPA	TOXCAST	ATG_NFI_CIS	NFIC	nuclear factor I/C (CCAAT-binding transcription factor)
EPA	TOXCAST	ATG_NFI_CIS	NFIX	nuclear factor I/X (CCAAT-binding transcription factor)
EPA	TOXCAST	ATG_NRF1_CIS	NRF1	nuclear respiratory factor 1
EPA	TOXCAST	ATG_NRF2_ARE_CIS	NFE2L2	nuclear factor (erythroid-derived 2)-like 2
EPA	TOXCAST	ATG_NURR1_TRANS	NR4A2	nuclear receptor subfamily 4, group A, member 2 (NGFI-B/nur77 beta-type transcription factor homolog)
EPA	TOXCAST	ATG_Oct_MLP_CIS	POU2F1	POU class 2 homeobox 1
EPA	TOXCAST	ATG_p53_CIS	TP53	tumor protein p53
EPA	TOXCAST	ATG_Pax6_CIS	PAX6	paired box 6
EPA	TOXCAST	ATG_PBREM_CIS	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
EPA	TOXCAST	ATG_PBREM_CIS	NR1I3	nuclear receptor subfamily 1, group I, member 3 (Constitutive Androstane Receptor)
EPA	TOXCAST	ATG_PPARa_TRANS	PPARA	peroxisome proliferator-activated receptor alpha
EPA	TOXCAST	ATG_PPARd_TRANS	PPARD	peroxisome proliferator-activated receptor delta
EPA	TOXCAST	ATG_PPARg_TRANS	PPARG	peroxisome proliferator-activated receptor gamma
EPA	TOXCAST	ATG_PPRE_CIS	PPARA	peroxisome proliferator-activated receptor alpha
EPA	TOXCAST	ATG_PPRE_CIS	PPARD	peroxisome proliferator-activated receptor delta
EPA	TOXCAST	ATG_PPRE_CIS	PPARG	peroxisome proliferator-activated receptor gamma
EPA	TOXCAST	ATG_PXR_TRANS	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
EPA	TOXCAST	ATG_PXRE_CIS	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
EPA	TOXCAST	ATG_RARa_TRANS	RARA	retinoic acid receptor, alpha

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	ATG_RARb_TRANS	RARB	retinoic acid receptor, beta
EPA	TOXCAST	ATG_RARg_TRANS	RARG	retinoic acid receptor, gamma
EPA	TOXCAST	ATG_RORb_TRANS	RORB	RAR-related orphan receptor B
EPA	TOXCAST	ATG_RORE_CIS	RARA	retinoic acid receptor, alpha
EPA	TOXCAST	ATG_RORg_TRANS	RORC	RAR-related orphan receptor C
EPA	TOXCAST	ATG_RXRa_TRANS	RXRA	retinoid X receptor, alpha
EPA	TOXCAST	ATG_RXRb_TRANS	RXRB	retinoid X receptor, beta
EPA	TOXCAST	ATG_Sox_CIS	SOX1	SRY (sex determining region Y)-box 1
EPA	TOXCAST	ATG_Sp1_CIS	SP1	Sp1 transcription factor
EPA	TOXCAST	ATG_SREBP_CIS	SREBF1	sterol regulatory element binding transcription factor 1
EPA	TOXCAST	ATG_STAT3_CIS	STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)
EPA	TOXCAST	ATG_TCF_b_cat_CIS	CTNNB1	catenin (cadherin-associated protein), beta 1, 88kDa
EPA	TOXCAST	ATG_TGFb_CIS	TGFB1	transforming growth factor, beta 1
EPA	TOXCAST	ATG_THRa1_TRANS	THRA	thyroid hormone receptor, alpha (erythroblastic leukemia viral (v-erb-a) oncogene homolog, avian)
EPA	TOXCAST	ATG_VDR_TRANS	VDR	vitamin D (1,25- dihydroxyvitamin D3) receptor
EPA	TOXCAST	ATG_VDRE_CIS	VDR	vitamin D (1,25- dihydroxyvitamin D3) receptor
EPA	TOXCAST	ATG_Xbp1_CIS	XBP1	X-box binding protein 1
EPA	TOXCAST	BSK_3C_Eselectin_down	SELE	selectin E
EPA	TOXCAST	BSK_3C_Eselectin_up	SELE	selectin E
EPA	TOXCAST	BSK_3C_hLADR_down	HLA-DRA	major histocompatibility complex, class II, DR alpha
EPA	TOXCAST	BSK_3C_hLADR_up	HLA-DRA	major histocompatibility complex, class II, DR alpha
EPA	TOXCAST	BSK_3C_ICAM1_down	ICAM1	intercellular adhesion molecule 1
EPA	TOXCAST	BSK_3C_ICAM1_up	ICAM1	intercellular adhesion molecule 1
EPA	TOXCAST	BSK_3C_IL8_down	IL8	interleukin 8
EPA	TOXCAST	BSK_3C_IL8_up	IL8	interleukin 8
EPA	TOXCAST	BSK_3C_MCP1_down	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_3C_MCP1_up	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_3C_MIG_down	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_3C_MIG_up	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_3C_Thrombomodulin_down	THBD	thrombomodulin
EPA	TOXCAST	BSK_3C_Thrombomodulin_up	THBD	thrombomodulin

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	BSK_3C_TissueFactor_ down	F3	coagulation factor III (thromboplastin, tissue factor)
EPA	TOXCAST	BSK_3C_TissueFactor_up	F3	coagulation factor III (thromboplastin, tissue factor)
EPA	TOXCAST	BSK_3C_uPAR_down	PLAUR	plasminogen activator, urokinase receptor
EPA	TOXCAST	BSK_3C_uPAR_up	PLAUR	plasminogen activator, urokinase receptor
EPA	TOXCAST	BSK_3C_VCAM1_down	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	BSK_3C_VCAM1_up	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	BSK_4H_Eotaxin3_down	CCL26	chemokine (C-C motif) ligand 26
EPA	TOXCAST	BSK_4H_Eotaxin3_up	CCL26	chemokine (C-C motif) ligand 26
EPA	TOXCAST	BSK_4H_MCP1_down	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_4H_MCP1_up	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_4H_Pselectin_down	SELP	selectin P (granule membrane protein 140kDa, antigen CD62)
EPA	TOXCAST	BSK_4H_Pselectin_up	SELP	selectin P (granule membrane protein 140kDa, antigen CD62)
EPA	TOXCAST	BSK_4H_uPAR_down	PLAUR	plasminogen activator, urokinase receptor
EPA	TOXCAST	BSK_4H_uPAR_up	PLAUR	plasminogen activator, urokinase receptor
EPA	TOXCAST	BSK_4H_VCAM1_down	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	BSK_4H_VCAM1_up	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	BSK_4H_VEGFRII_down	KDR	kinase insert domain receptor (a type III receptor tyrosine kinase)
EPA	TOXCAST	BSK_4H_VEGFRII_up	KDR	kinase insert domain receptor (a type III receptor tyrosine kinase)
EPA	TOXCAST	BSK_BE3C_hLADR_down	HLA-DRA	major histocompatibility complex, class II, DR alpha
EPA	TOXCAST	BSK_BE3C_hLADR_up	HLA-DRA	major histocompatibility complex, class II, DR alpha
EPA	TOXCAST	BSK_BE3C_IL1a_down	IL1A	interleukin 1, alpha
EPA	TOXCAST	BSK_BE3C_IL1a_up	IL1A	interleukin 1, alpha
EPA	TOXCAST	BSK_BE3C_IP10_down	CXCL10	chemokine (C-X-C motif) ligand 10
EPA	TOXCAST	BSK_BE3C_IP10_up	CXCL10	chemokine (C-X-C motif) ligand 10
EPA	TOXCAST	BSK_BE3C_MIG_down	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_BE3C_MIG_up	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_BE3C_MMP1_down	MMP1	matrix metallopeptidase 1 (interstitial collagenase)
EPA	TOXCAST	BSK_BE3C_MMP1_up	MMP1	matrix metallopeptidase 1 (interstitial collagenase)
EPA	TOXCAST	BSK_BE3C_PAI1_down	SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1
EPA	TOXCAST	BSK_BE3C_PAI1_up	SERPINE2	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2
EPA	TOXCAST	BSK_BE3C_TGFb1_down	TGFB1	transforming growth factor, beta 1

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	BSK_BE3C_TGFb1_up	TGFB1	transforming growth factor, beta 1
EPA	TOXCAST	BSK_BE3C_tPA_down	PLAT	plasminogen activator, tissue
EPA	TOXCAST	BSK_BE3C_tPA_up	PLAT	plasminogen activator, tissue
EPA	TOXCAST	BSK_BE3C_uPA_down	PLAU	plasminogen activator, urokinase
EPA	TOXCAST	BSK_BE3C_uPA_up	PLAU	plasminogen activator, urokinase
EPA	TOXCAST	BSK_BE3C_uPAR_down	PLAUR	plasminogen activator, urokinase receptor
EPA	TOXCAST	BSK_BE3C_uPAR_up	PLAUR	plasminogen activator, urokinase receptor
EPA	TOXCAST	BSK_hDFCGF_CollagenIII_ down	COL3A1	collagen, type III, alpha 1
EPA	TOXCAST	BSK_hDFCGF_CollagenIII_ up	COL3A1	collagen, type III, alpha 1
EPA	TOXCAST	BSK_hDFCGF_EGFR_down	EGFR	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)
EPA	TOXCAST	BSK_hDFCGF_EGFR_up	EGFR	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)
EPA	TOXCAST	BSK_hDFCGF_IL8_down	IL8	interleukin 8
EPA	TOXCAST	BSK_hDFCGF_IL8_up	IL9	interleukin 9
EPA	TOXCAST	BSK_hDFCGF_IP10_down	CXCL10	chemokine (C-X-C motif) ligand 10
EPA	TOXCAST	BSK_hDFCGF_IP10_up	CXCL11	chemokine (C-X-C motif) ligand 11
EPA	TOXCAST	BSK_hDFCGF_MCSF_down	CSF1	colony stimulating factor 1 (macrophage)
EPA	TOXCAST	BSK_hDFCGF_MCSF_up	CSF1	colony stimulating factor 1 (macrophage)
EPA	TOXCAST	BSK_hDFCGF_MIG_down	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_hDFCGF_MIG_up	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_hDFCGF_MMP1_ down	MMP1	matrix metallopeptidase 1 (interstitial collagenase)
EPA	TOXCAST	BSK_hDFCGF_MMP1_up	MMP1	matrix metallopeptidase 1 (interstitial collagenase)
EPA	TOXCAST	BSK_hDFCGF_PAI1_down	SERPINE3	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 3
EPA	TOXCAST	BSK_hDFCGF_PAI1_up	SERPINE3	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 3
EPA	TOXCAST	BSK_hDFCGF_TIMP1_ down	TIMP1	TIMP metallopeptidase inhibitor 1
EPA	TOXCAST	BSK_hDFCGF_TIMP1_up	TIMP1	TIMP metallopeptidase inhibitor 1
EPA	TOXCAST	BSK_hDFCGF_VCAM1_ down	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	BSK_hDFCGF_VCAM1_up	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	BSK_KF3CT_ICAM1_down	ICAM1	intercellular adhesion molecule 1
EPA	TOXCAST	BSK_KF3CT_ICAM1_up	ICAM1	intercellular adhesion molecule 1

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	BSK_KF3CT_IL1a_down	IL1A	interleukin 1, alpha
EPA	TOXCAST	BSK_KF3CT_IL1a_up	IL1A	interleukin 1, alpha
EPA	TOXCAST	BSK_KF3CT_IP10_down	CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)
EPA	TOXCAST	BSK_KF3CT_IP10_up	CXCL13	chemokine (C-X-C motif) ligand 13
EPA	TOXCAST	BSK_KF3CT_MCP1_down	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_KF3CT_MCP1_up	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_KF3CT_MMP9_down	MMP9	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)
EPA	TOXCAST	BSK_KF3CT_MMP9_up	MMP9	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)
EPA	TOXCAST	BSK_KF3CT_TGFb1_down	TGFB2	transforming growth factor, beta 2
EPA	TOXCAST	BSK_KF3CT_TGFb1_up	TGFB3	transforming growth factor, beta 3
EPA	TOXCAST	BSK_KF3CT_TIMP2_down	TIMP2	TIMP metallopeptidase inhibitor 2
EPA	TOXCAST	BSK_KF3CT_TIMP2_up	TIMP3	TIMP metallopeptidase inhibitor 3
EPA	TOXCAST	BSK_KF3CT_uPA_down	PLAU	plasminogen activator, urokinase
EPA	TOXCAST	BSK_KF3CT_uPA_up	PLAU	plasminogen activator, urokinase
EPA	TOXCAST	BSK_LPS_CD40_down	CD40	CD40 molecule, TNF receptor superfamily member 5
EPA	TOXCAST	BSK_LPS_CD40_up	CD40	CD40 molecule, TNF receptor superfamily member 5
EPA	TOXCAST	BSK_LPS_Eselectin_down	SELE	selectin E
EPA	TOXCAST	BSK_LPS_Eselectin_up	SELE	selectin E
EPA	TOXCAST	BSK_LPS_IL1a_down	IL1A	interleukin 1, alpha
EPA	TOXCAST	BSK_LPS_IL1a_up	IL1A	interleukin 1, alpha
EPA	TOXCAST	BSK_LPS_IL8_down	IL10	interleukin 10
EPA	TOXCAST	BSK_LPS_IL8_up	IL11	interleukin 11
EPA	TOXCAST	BSK_LPS_MCSF_down	CSF1	colony stimulating factor 1 (macrophage)
EPA	TOXCAST	BSK_LPS_MCSF_up	CSF1	colony stimulating factor 1 (macrophage)
EPA	TOXCAST	BSK_LPS_MPC1_down	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_LPS_MPC1_up	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_LPS_PGE2_down	PTGER2	prostaglandin E receptor 2 (subtype EP2), 53kDa
EPA	TOXCAST	BSK_LPS_PGE2_up	PTGER2	prostaglandin E receptor 2 (subtype EP2), 53kDa
EPA	TOXCAST	BSK_LPS_TissueFactor_ down	F3	coagulation factor III (thromboplastin, tissue factor)
EPA	TOXCAST	BSK_LPS_TissueFactor_ up	F3	coagulation factor III (thromboplastin, tissue factor)
EPA	TOXCAST	BSK_LPS_TNFa_down	TNF	tumor necrosis factor (TNF superfamily, member 2)
EPA	TOXCAST	BSK_LPS_TNFa_up	TNF	tumor necrosis factor (TNF superfamily, member 2)

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	BSK_LPS_VCAM1_down	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	BSK_LPS_VCAM1_up	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	BSK_SAg_CD38_down	CD38	CD38 molecule
EPA	TOXCAST	BSK_SAg_CD38_up	CD38	CD38 molecule
EPA	TOXCAST	BSK_SAg_CD40_down	CD40	CD40 molecule, TNF receptor superfamily member 5
EPA	TOXCAST	BSK_SAg_CD40_up	CD40	CD40 molecule, TNF receptor superfamily member 5
EPA	TOXCAST	BSK_SAg_CD69_down	CD69	CD69 molecule
EPA	TOXCAST	BSK_SAg_CD69_up	CD69	CD69 molecule
EPA	TOXCAST	BSK_SAg_Eselectin_ down	SELE	selectin E
EPA	TOXCAST	BSK_SAg_Eselectin_up	SELE	selectin E
EPA	TOXCAST	BSK_SAg_IL8_down	IL12B	interleukin 12B
EPA	TOXCAST	BSK_SAg_IL8_up	IL13	interleukin 13
EPA	TOXCAST	BSK_SAg_MCP1_down	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_SAg_MCP1_up	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_SAg_MIG_down	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_SAg_MIG_up	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_SM3C_HLADR_ down	HLA-DRA	major histocompatibility complex, class II, DR alpha
EPA	TOXCAST	BSK_SM3C_HLADR_up	HLA-DRA	major histocompatibility complex, class II, DR alpha
EPA	TOXCAST	BSK_SM3C_IL_6_down	IL6	interleukin 6 (interferon, beta 2)
EPA	TOXCAST	BSK_SM3C_IL_6_up	IL6	interleukin 6 (interferon, beta 2)
EPA	TOXCAST	BSK_SM3C_IL_8_down	TXLNA	taxilin alpha
EPA	TOXCAST	BSK_SM3C_IL_8_up	IL15	interleukin 15
EPA	TOXCAST	BSK_SM3C_LDLR_down	LDLR	low density lipoprotein receptor
EPA	TOXCAST	BSK_SM3C_LDLR_up	LDLR	low density lipoprotein receptor
EPA	TOXCAST	BSK_SM3C_MCP1_down	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_SM3C_MCP1_up	CCL2	chemokine (C-C motif) ligand 2
EPA	TOXCAST	BSK_SM3C_MCSF_down	CSF1	colony stimulating factor 1 (macrophage)
EPA	TOXCAST	BSK_SM3C_MCSF_up	CSF1	colony stimulating factor 1 (macrophage)
EPA	TOXCAST	BSK_SM3C_MIG_down	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_SM3C_MIG_up	CXCL9	chemokine (C-X-C motif) ligand 9
EPA	TOXCAST	BSK_SM3C_SAA_down	SAA1	serum amyloid A1
EPA	TOXCAST	BSK_SM3C_SAA_up	SAA1	serum amyloid A2
EPA	TOXCAST	BSK_SM3C_Thrombomodulin_down	THBD	thrombomodulin

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	BSK_SM3C_Thrombomodulin_up	THBD	thrombomodulin
EPA	TOXCAST	BSK_SM3C_TissueFactor_down	F3	coagulation factor III (thromboplastin, tissue factor)
EPA	TOXCAST	BSK_SM3C_TissueFactor_up	F3	coagulation factor III (thromboplastin, tissue factor)
EPA	TOXCAST	BSK_SM3C_uPAR_down	PLAUR	plasminogen activator, urokinase receptor
EPA	TOXCAST	BSK_SM3C_uPAR_up	PLAUR	plasminogen activator, urokinase receptor
EPA	TOXCAST	BSK_SM3C_VCAM_1_down	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	BSK_SM3C_VCAM_1_up	VCAM1	vascular cell adhesion molecule 1
EPA	TOXCAST	CLZD_ABCB1	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1
EPA	TOXCAST	CLZD_ABCB11	ABCB11	ATP-binding cassette, sub-family B (MDR/TAP), member 11
EPA	TOXCAST	CLZD_ABCG2	ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2
EPA	TOXCAST	CLZD_ACTIN	ACTA1	actin, alpha 1, skeletal muscle
EPA	TOXCAST	CLZD_CYP1A1	CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1
EPA	TOXCAST	CLZD_CYP1A2	CYP1A2	cytochrome P450, family 1, subfamily A, polypeptide 2
EPA	TOXCAST	CLZD_CYP2B6	Cyp2B6	cytochrome P450, family 2, subfamily B, polypeptide 6
EPA	TOXCAST	CLZD_CYP2C19	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19
EPA	TOXCAST	CLZD_CYP2C9	CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9
EPA	TOXCAST	CLZD_CYP2C9	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
EPA	TOXCAST	CLZD_CYP2C9	NR1I3	nuclear receptor subfamily 1, group I, member 3 (Constitutive Androstane Receptor)
EPA	TOXCAST	CLZD_CYP2C9	NR1I3	nuclear receptor subfamily 1, group I, member 3 (Constitutive Androstane Receptor)
EPA	TOXCAST	CLZD_CYP3A4	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4
EPA	TOXCAST	CLZD_CYP3A4	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
EPA	TOXCAST	CLZD_CYP3A4	NR1I3	nuclear receptor subfamily 1, group I, member 3 (Constitutive Androstane Receptor)
EPA	TOXCAST	CLZD_GAPDH	GAPDH	glyceraldehyde-3-phosphate dehydrogenase-like 6; hypothetical protein LOC100133042; glyceraldehyde-3-phosphate dehydrogenase
EPA	TOXCAST	CLZD_GSTA2	GSTA2	glutathione S-transferase alpha 2
EPA	TOXCAST	CLZD_HMGCS2	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)
EPA	TOXCAST	CLZD_SLCO1B1	SLCO1B1	solute carrier organic anion transporter family, member 1B1
EPA	TOXCAST	CLZD_SULT2A1	SULT2A1	sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone (DHEA)-preferring, member 1

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	CLZD_UGT1A1	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
EPA	TOXCAST	CLZD_UGT1A1	NR1I3	nuclear receptor subfamily 1, group I, member 3 (Constitutive Androstane Receptor)
EPA	TOXCAST	DNADamage	TP53	tumor protein p53
EPA	TOXCAST	MicroTubule	TUBA1A	tubulin, alpha 1a
EPA	TOXCAST	NCGC_AR_Agonist	AR	androgen receptor
EPA	TOXCAST	NCGC_AR_Antagonist	AR	androgen receptor
EPA	TOXCAST	NCGC_ERalpha_Agonist	ESR1	estrogen receptor 1
EPA	TOXCAST	NCGC_ERalpha_ Antagonist	ESR1	estrogen receptor 1
EPA	TOXCAST	NCGC_FXR_Agonist	NR1H4	nuclear receptor subfamily 1, group H, member 4 (Farnesoid X Receptor)
EPA	TOXCAST	NCGC_GR_Agonist	NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
EPA	TOXCAST	NCGC_LXR_Agonist	NR1H2	nuclear receptor subfamily 1, group H, member 2(Liver X receptor beta)
EPA	TOXCAST	NCGC_p53	TP53	tumor protein p53
EPA	TOXCAST	NCGC_PPARa_Agonist	PPARA	peroxisome proliferator-activated receptor alpha
EPA	TOXCAST	NCGC_PPARd_Agonist	PPARD	peroxisome proliferator-activated receptor delta
EPA	TOXCAST	NCGC_PPARg_Agonist	PPARG	peroxisome proliferator-activated receptor gamma
EPA	TOXCAST	NCGC_PXR_Agonist_ human	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
EPA	TOXCAST	NCGC_PXR_Agonist_rat	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
EPA	TOXCAST	NCGC_RXRa_Agonist	RXRA	retinoid X receptor, alpha
EPA	TOXCAST	NCGC_TRbeta_Agonist	THRB	thyroid hormone receptor, beta (erythroblastic leukemia viral (v-erb-a) oncogene homolog 2, avian)
EPA	TOXCAST	NCGC_TRbeta_Antagonist	THRB	thyroid hormone receptor, beta (erythroblastic leukemia viral (v-erb-a) oncogene homolog 2, avian)
EPA	TOXCAST	NCGC_VDR_Agonist	VDR	vitamin D (1,25- dihydroxyvitamin D3) receptor
EPA	TOXCAST	NVS_ADME_hCYP1A1	CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1
EPA	TOXCAST	NVS_ADME_hCYP1A2	CYP1A2	cytochrome P450, family 1, subfamily A, polypeptide 2
EPA	TOXCAST	NVS_ADME_hCYP1B1	CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1
EPA	TOXCAST	NVS_ADME_hCYP2A6	CYP2A6	cytochrome P450, family 2, subfamily A, polypeptide 6
EPA	TOXCAST	NVS_ADME_hCYP2B6	Cyp2B6	cytochrome P450, family 2, subfamily B, polypeptide 6
EPA	TOXCAST	NVS_ADME_hCYP2C18	CYP2C18	cytochrome P450, family 2, subfamily C, polypeptide 18
EPA	TOXCAST	NVS_ADME_hCYP2C19	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	NVS_ADME_hCYP2C19_ Activator	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19
EPA	TOXCAST	NVS_ADME_hCYP2C8	CYP2C8	cytochrome P450, family 2, subfamily C, polypeptide 8
EPA	TOXCAST	NVS_ADME_hCYP2C9	CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9
EPA	TOXCAST	NVS_ADME_hCYP2D6	CYP2D6	cytochrome P450, family 2, subfamily D, polypeptide 6
EPA	TOXCAST	NVS_ADME_hCYP2E1	CYP2E1	cytochrome P450, family 2, subfamily E, polypeptide 1
EPA	TOXCAST	NVS_ADME_hCYP2J2	CYP2J2	cytochrome P450, family 2, subfamily J, polypeptide 2
EPA	TOXCAST	NVS_ADME_hCYP3A4	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4
EPA	TOXCAST	NVS_ADME_hCYP3A5	CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5
EPA	TOXCAST	NVS_ADME_hCYP4F12	CYP4F12	similar to cytochrome P450, family 4, subfamily F, polypeptide 12; cytochrome P450, family 4, subfamily F, polypeptide 12
EPA	TOXCAST	NVS_ADME_hCYP4F12_ Activator	CYP4F12	similar to cytochrome P450, family 4, subfamily F, polypeptide 12; cytochrome P450, family 4, subfamily F, polypeptide 12
EPA	TOXCAST	NVS_ADME_hCYPC19	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19
EPA	TOXCAST	NVS_ADME_rCYP1A1	CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1
EPA	TOXCAST	NVS_ADME_rCYP1A2	CYP1A2	cytochrome P450, family 1, subfamily A, polypeptide 2
EPA	TOXCAST	NVS_ADME_rCYP2A1	Cyp2A1	cytochrome P450, family 2, subfamily a, polypeptide 1
EPA	TOXCAST	NVS_ADME_rCYP2A2	Cyp2A2	cytochrome P450, family 2, subfamily a, polypeptide 2
EPA	TOXCAST	NVS_ADME_rCYP2B1	Cyp2B1	cytochrome P450, family 2, subfamily b, polypeptide 1
EPA	TOXCAST	NVS_ADME_rCYP2C11	Cyp2c11	cytochrome P450, subfamily 2, polypeptide 11
EPA	TOXCAST	NVS_ADME_rCYP2C12	Cyp2c12	cytochrome P450, subfamily 2, polypeptide 12
EPA	TOXCAST	NVS_ADME_rCYP2C13	Cyp2c13	cytochrome P450, subfamily 2, polypeptide 13
EPA	TOXCAST	NVS_ADME_rCYP2C6	Cyp2c6	cytochrome P450, family 2, subfamily c, polypeptide 6
EPA	TOXCAST	NVS_ADME_rCYP2D1	Cyp2d1	cytochrome P450, family 2, subfamily d, polypeptide 1
EPA	TOXCAST	NVS_ADME_rCYP2D2	Cyp2d2	cytochrome P450, family 2, subfamily d, polypeptide 2
EPA	TOXCAST	NVS_ADME_rCYP2E1	Cyp2e1	cytochrome P450, family 2, subfamily E, polypeptide 1
EPA	TOXCAST	NVS_ADME_rCYP3A1	Cyp3a1	cytochrome P450, family 3, subfamily a, polypeptide 1
EPA	TOXCAST	NVS_ADME_rCYP3A2	Cyp3a2	cytochrome P450, family 3, subfamily a, polypeptide 2
EPA	TOXCAST	NVS_ENZ_hAbl	ABL1	c-abl oncogene 1, receptor tyrosine kinase
EPA	TOXCAST	NVS_ENZ_hAbl_ Activator	ABL1	c-abl oncogene 1, receptor tyrosine kinase
EPA	TOXCAST	NVS_ENZ_hAChE	ACHE	acetylcholinesterase (Yt blood group)
EPA	TOXCAST	NVS_ENZ_hAurA	AURKA	aurora kinase A; aurora kinase A pseudogene 1
EPA	TOXCAST	NVS_ENZ_hBACE	BACE1	beta-site APP-cleaving enzyme 1
EPA	TOXCAST	NVS_ENZ_hBTK	BTK	Bruton agammaglobulinemia tyrosine kinase
EPA	TOXCAST	NVS_ENZ_hBTK_ Activator	BTK	Bruton agammaglobulinemia tyrosine kinase

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	NVS_ENZ_hCASP1	CASP1	caspase 1, apoptosis-related cysteine peptidase (interleukin 1, beta, convertase)
EPA	TOXCAST	NVS_ENZ_hCASP10	CASP10	caspase 10, apoptosis-related cysteine peptidase
EPA	TOXCAST	NVS_ENZ_hCASP2	CASP2	caspase 2, apoptosis-related cysteine peptidase
EPA	TOXCAST	NVS_ENZ_hCASP3	CASP3	caspase 3, apoptosis-related cysteine peptidase
EPA	TOXCAST	NVS_ENZ_hCASP4	CASP4	caspase 4, apoptosis-related cysteine peptidase
EPA	TOXCAST	NVS_ENZ_hCASP5	CASP5	caspase 5, apoptosis-related cysteine peptidase
EPA	TOXCAST	NVS_ENZ_hCASP8	CASP8	caspase 8, apoptosis-related cysteine peptidase
EPA	TOXCAST	NVS_ENZ_hCD45	PTPRC	protein tyrosine phosphatase, receptor type, C
EPA	TOXCAST	NVS_ENZ_hCHK1	CHEK1	CHK1 checkpoint homolog (S. pombe)
EPA	TOXCAST	NVS_ENZ_hCK1D	CSNK1D	casein kinase 1, delta
EPA	TOXCAST	NVS_ENZ_hCK2a2b2	CSNK2A1	casein kinase 2, alpha 1 polypeptide pseudogene; casein kinase 2, alpha 1 polypeptide
EPA	TOXCAST	NVS_ENZ_hElastase	ELANE	elastase, neutrophil expressed
EPA	TOXCAST	NVS_ENZ_hES	BCHE	butyrylcholinesterase
EPA	TOXCAST	NVS_ENZ_hFyn	FYN	FYN oncogene related to SRC, FGR, YES
EPA	TOXCAST	NVS_ENZ_hFyn_ Activator	FYN	FYN oncogene related to SRC, FGR, YES
EPA	TOXCAST	NVS_ENZ_hGSK3b	GSK3B	glycogen synthase kinase 3 beta
EPA	TOXCAST	NVS_ENZ_hIKKa	CHUK	conserved helix-loop-helix ubiquitous kinase
EPA	TOXCAST	NVS_ENZ_hInsR	INSR	insulin receptor
EPA	TOXCAST	NVS_ENZ_hInsR_ Activator	INSR	insulin receptor
EPA	TOXCAST	NVS_ENZ_hIRAK4	IRAK4	interleukin-1 receptor-associated kinase 4
EPA	TOXCAST	NVS_ENZ_hLck	LCK	lymphocyte-specific protein tyrosine kinase
EPA	TOXCAST	NVS_ENZ_hLMPTPA	ACP1	acid phosphatase 1, soluble
EPA	TOXCAST	NVS_ENZ_hLynA	LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog
EPA	TOXCAST	NVS_ENZ_hLynA_ Activator	LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog
EPA	TOXCAST	NVS_ENZ_hMAPK3	MAPK3	hypothetical LOC100271831; mitogen-activated protein kinase 3
EPA	TOXCAST	NVS_ENZ_hMAPKAPK2	MAPKAPK2	mitogen-activated protein kinase-activated protein kinase 2
EPA	TOXCAST	NVS_ENZ_hMAPKAPK5	MAPKAPK5	mitogen-activated protein kinase-activated protein kinase 5
EPA	TOXCAST	NVS_ENZ_hMARK1	MARK1	MAP/microtubule affinity-regulating kinase 1
EPA	TOXCAST	NVS_ENZ_hMet	MET	met proto-oncogene (hepatocyte growth factor receptor)
EPA	TOXCAST	NVS_ENZ_hMMP1	MMP1	matrix metallopeptidase 1 (interstitial collagenase)
EPA	TOXCAST	NVS_ENZ_hMMP13	MMP13	matrix metallopeptidase 13 (collagenase 3)
EPA	TOXCAST	NVS_ENZ_hMMP2	MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	NVS_ENZ_hMMP3	MMP3	matrix metallopeptidase 3 (stromelysin 1, progelatinase)
EPA	TOXCAST	NVS_ENZ_hMMP7	MMP7	matrix metallopeptidase 7 (matrilysin, uterine)
EPA	TOXCAST	NVS_ENZ_hMMP9	MMP9	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)
EPA	TOXCAST	NVS_ENZ_hMsk1	RPS6KA5	ribosomal protein S6 kinase, 90kDa, polypeptide 5
EPA	TOXCAST	NVS_ENZ_hPKA	PRKACA	protein kinase, cAMP-dependent, catalytic, alpha
EPA	TOXCAST	NVS_ENZ_hPKBa	AKT1	v-akt murine thymoma viral oncogene homolog 1
EPA	TOXCAST	NVS_ENZ_hPKBb	AKT2	v-akt murine thymoma viral oncogene homolog 2
EPA	TOXCAST	NVS_ENZ_hPKCz	PRKCZ	protein kinase C, zeta
EPA	TOXCAST	NVS_ENZ_hPKD2	PRKD2	protein kinase D2
EPA	TOXCAST	NVS_ENZ_hPP1a	PPP1CA	protein phosphatase 1, catalytic subunit, alpha isoform
EPA	TOXCAST	NVS_ENZ_hPP2A	PPP2CA	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform
EPA	TOXCAST	NVS_ENZ_hPP2Ca	PPM1A	protein phosphatase 1A (formerly 2C), magnesium-dependent, alpha isoform
EPA	TOXCAST	NVS_ENZ_hPP2Ca	PPM1A	protein phosphatase 1A (formerly 2C), magnesium-dependent, alpha isoform
EPA	TOXCAST	NVS_ENZ_hPP2Ca_ Activator	PPM1A	protein phosphatase 1A (formerly 2C), magnesium-dependent, alpha isoform
EPA	TOXCAST	NVS_ENZ_hPP2Ca_ Activator	PPM1A	protein phosphatase 1A (formerly 2C), magnesium-dependent, alpha isoform
EPA	TOXCAST	NVS_ENZ_hPPVHR	DUSP3	dual specificity phosphatase 3
EPA	TOXCAST	NVS_ENZ_hPTP1b	PTPN1	protein tyrosine phosphatase, non-receptor type 1
EPA	TOXCAST	NVS_ENZ_hPTPb	PTPRB	protein tyrosine phosphatase, receptor type, B
EPA	TOXCAST	NVS_ENZ_hPTPBAS	PTPN13	protein tyrosine phosphatase, non-receptor type 13 (APO-1/CD95 (Fas)-associated phosphatase)
EPA	TOXCAST	NVS_ENZ_hPTPD2	PTPN14	protein tyrosine phosphatase, non-receptor type 14
EPA	TOXCAST	NVS_ENZ_hPTPLAR	PTPRF	protein tyrosine phosphatase, receptor type, F
EPA	TOXCAST	NVS_ENZ_hPTPMEG1	PTPN4	protein tyrosine phosphatase, non-receptor type 4 (megakaryocyte)
EPA	TOXCAST	NVS_ENZ_hPTPMEG2	PTPN9	protein tyrosine phosphatase, non-receptor type 9
EPA	TOXCAST	NVS_ENZ_hPTPMU	PTPRM	protein tyrosine phosphatase, receptor type, M
EPA	TOXCAST	NVS_ENZ_hPTPPEST	PTPN12	protein tyrosine phosphatase, non-receptor type 12
EPA	TOXCAST	NVS_ENZ_hPTPSHP1	PTPN6	protein tyrosine phosphatase, non-receptor type 6
EPA	TOXCAST	NVS_ENZ_hPTPSHP2	PTPN11	protein tyrosine phosphatase, non-receptor type 11; similar to protein tyrosine phosphatase, non-receptor type 11
EPA	TOXCAST	NVS_ENZ_hPTPT	PTPRT	protein tyrosine phosphatase, receptor type, T
EPA	TOXCAST	NVS_ENZ_hSGK1	SGK1	serum/glucocorticoid regulated kinase 1

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	NVS_ENZ_hSRC	SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)
EPA	TOXCAST	NVS_ENZ_hSRC_ Activator	SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)
EPA	TOXCAST	NVS_ENZ_hTie2	TEK	TEK tyrosine kinase, endothelial
EPA	TOXCAST	NVS_ENZ_hZAP70	ZAP70	zeta-chain (TCR) associated protein kinase 70kDa
EPA	TOXCAST	NVS_ENZ_oCOX1	PTGS1	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)
EPA	TOXCAST	NVS_ENZ_oCOX2	PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)
EPA	TOXCAST	NVS_ENZ_pMTHFR	MTHFR	5,10-methylenetetrahydrofolate reductase (NADPH)
EPA	TOXCAST	NVS_ENZ_rabI2C	MAOA	monoamine oxidase A
EPA	TOXCAST	NVS_ENZ_rabI2C	MAOB	monoamine oxidase B
EPA	TOXCAST	NVS_ENZ_rACFSKBinding	Adcy5	adenylate cyclase 5
EPA	TOXCAST	NVS_ENZ_rAChE	Ache	acetylcholinesterase (Yt blood group)
EPA	TOXCAST	NVS_ENZ_rCNOS	Nos1	nitric oxide synthase 1 (neuronal)
EPA	TOXCAST	NVS_ENZ_rCOMT	COMT	catechol-O-methyltransferase
EPA	TOXCAST	NVS_ENZ_rMAOAC	Maoa	monoamine oxidase A
EPA	TOXCAST	NVS_ENZ_rMAOAP	Maoa	monoamine oxidase A
EPA	TOXCAST	NVS_ENZ_rMAOBC	Maob	monoamine oxidase B
EPA	TOXCAST	NVS_ENZ_rMOABP	Maob	monoamine oxidase B
EPA	TOXCAST	NVS_GPCR_bAdoRNon Selective	ADRA1A	adrenergic, alpha-1A-, receptor
EPA	TOXCAST	NVS_GPCR_bAT2	AGTR2	angiotensin II receptor, type 2
EPA	TOXCAST	NVS_GPCR_bDR_Non_ Selective	DRD1	dopamine receptor D1
EPA	TOXCAST	NVS_GPCR_bH1	HRH1	histamine receptor H1
EPA	TOXCAST	NVS_GPCR_bNPYNon_ Selective	NPY	neuropeptide Y
EPA	TOXCAST	NVS_GPCR_g5HT4	HTR4	5-hydroxytryptamine (serotonin) receptor 4
EPA	TOXCAST	NVS_GPCR_gANPA	Nppa	natriuretic peptide precursor A
EPA	TOXCAST	NVS_GPCR_gANPA	NPR1	natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)
EPA	TOXCAST	NVS_GPCR_gBK2	Bdkrb2	bradykinin receptor B2
EPA	TOXCAST	NVS_GPCR_gH2	Hrh2	histamine receptor H2
EPA	TOXCAST	NVS_GPCR_gLTB4	LTB4R	leukotriene B4 receptor
EPA	TOXCAST	NVS_GPCR_gLTD4	CYSLTR1	cysteinyl leukotriene receptor 1

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	NVS_GPCR_gMNon SelectivePeripheral	CHRM1	cholinergic receptor, muscarinic 1
EPA	TOXCAST	NVS_GPCR_gMNon SelectivePeripheral	CHRM2	cholinergic receptor, muscarinic 2
EPA	TOXCAST	NVS_GPCR_gMNon SelectivePeripheral	CHRM3	cholinergic receptor, muscarinic 3
EPA	TOXCAST	NVS_GPCR_gOpiateK	OPRK1	opioid receptor, kappa 1
EPA	TOXCAST	NVS_GPCR_h5HT2A	HTR2A	5-hydroxytryptamine (serotonin) receptor 2A
EPA	TOXCAST	NVS_GPCR_h5HT2C	HTR2C	5-hydroxytryptamine (serotonin) receptor 2C
EPA	TOXCAST	NVS_GPCR_h5HT5A	HTR5A	5-hydroxytryptamine (serotonin) receptor 5A
EPA	TOXCAST	NVS_GPCR_h5HT6	HTR6	5-hydroxytryptamine (serotonin) receptor 6
EPA	TOXCAST	NVS_GPCR_h5HT7	HTR7	5-hydroxytryptamine (serotonin) receptor 7 (adenylate cyclase-coupled)
EPA	TOXCAST	NVS_GPCR_hAdoRA1	ADORA1	adenosine A1 receptor
EPA	TOXCAST	NVS_GPCR_hAdoRA2a	ADORA2A	adenosine A2a receptor
EPA	TOXCAST	NVS_GPCR_hAdra2A	ADRA2A	adrenergic, alpha-2A-, receptor
EPA	TOXCAST	NVS_GPCR_hAdra2C	ADRA2C	adrenergic, alpha-2C-, receptor
EPA	TOXCAST	NVS_GPCR_hAdrb1	ADRB1	adrenergic, beta-1-, receptor
EPA	TOXCAST	NVS_GPCR_hAdrb2	ADRB2	adrenergic, beta-2-, receptor, surface
EPA	TOXCAST	NVS_GPCR_hAdrb3	ADRB3	adrenergic, beta-3-, receptor
EPA	TOXCAST	NVS_GPCR_hAT1	AGTR1	angiotensin II receptor, type 1
EPA	TOXCAST	NVS_GPCR_hC5a	C5AR1	complement component 5a receptor 1
EPA	TOXCAST	NVS_GPCR_hDRD1	DRD1	dopamine receptor D1
EPA	TOXCAST	NVS_GPCR_hDRD2s	DRD2	dopamine receptor D2
EPA	TOXCAST	NVS_GPCR_hDRD4.4	DRD4	dopamine receptor D4
EPA	TOXCAST	NVS_GPCR_hETA	EDNRA	endothelin receptor type A
EPA	TOXCAST	NVS_GPCR_hETB	EDNRB	endothelin receptor type B
EPA	TOXCAST	NVS_GPCR_hGalanin	Gal	galanin prepropeptide
EPA	TOXCAST	NVS_GPCR_hH1	HRH1	histamine receptor H1
EPA	TOXCAST	NVS_GPCR_hLTB4_BLT1	LTB4R	leukotriene B4 receptor
EPA	TOXCAST	NVS_GPCR_hM1	CHRM1	cholinergic receptor, muscarinic 1
EPA	TOXCAST	NVS_GPCR_hM2	CHRM2	cholinergic receptor, muscarinic 2
EPA	TOXCAST	NVS_GPCR_hM3	CHRM3	cholinergic receptor, muscarinic 3
EPA	TOXCAST	NVS_GPCR_hM4	CHRM4	cholinergic receptor, muscarinic 4
EPA	TOXCAST	NVS_GPCR_hM5	CHRM5	cholinergic receptor, muscarinic 5
EPA	TOXCAST	NVS_GPCR_hNK2	TACR2	tachykinin receptor 2

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	NVS_GPCR_hNPY1	NPY1R	neuropeptide Y receptor Y1
EPA	TOXCAST	NVS_GPCR_hNPY2	NPY2R	neuropeptide Y receptor Y2
EPA	TOXCAST	NVS_GPCR_hNTS	NTSR1	neurotensin receptor 1 (high affinity)
EPA	TOXCAST	NVS_GPCR_hOpiate_D2	DRD2	dopamine receptor D2
EPA	TOXCAST	NVS_GPCR_hOpiate_mu	OPRM1	opioid receptor, mu 1
EPA	TOXCAST	NVS_GPCR_hORL1	OPRL1	opiate receptor-like 1
EPA	TOXCAST	NVS_GPCR_hPY2	P2RY1	purinergic receptor P2Y, G-protein coupled, 1
EPA	TOXCAST	NVS_GPCR_hTXA2	TBXA2R	thromboxane A2 receptor
EPA	TOXCAST	NVS_GPCR_hV1A	AVPR1A	arginine vasopressin receptor 1A
EPA	TOXCAST	NVS_GPCR_mCCKA Peripheral	Cckar	cholecystokinin A receptor
EPA	TOXCAST	NVS_GPCR_mCKKB Central	Cckbr	cholecystokinin B receptor
EPA	TOXCAST	NVS_GPCR_r5HT1 NonSelective	Htr1a	5-hydroxytryptamine (serotonin) receptor 1A
EPA	TOXCAST	NVS_GPCR_r5HT NonSelective	Htr1a	5-hydroxytryptamine (serotonin) receptor 1A
EPA	TOXCAST	NVS_GPCR_r5HT NonSelective	Htr1b	5-hydroxytryptamine (serotonin) receptor 1B
EPA	TOXCAST	NVS_GPCR_r5HT NonSelective	Htr2a	5-hydroxytryptamine (serotonin) receptor 2A
EPA	TOXCAST	NVS_GPCR_r5HT NonSelective	Htr2c	5-hydroxytryptamine (serotonin) receptor 2C
EPA	TOXCAST	NVS_GPCR_rabPAF	PTAFR	platelet-activating factor receptor
EPA	TOXCAST	NVS_GPCR_rAdra1A	Adra1a	adrenergic, alpha-1A-, receptor
EPA	TOXCAST	NVS_GPCR_rAdra1B	Adra1b	adrenergic, alpha-1B-, receptor
EPA	TOXCAST	NVS_GPCR_rAdra1 NonSelective	Adra1a	adrenergic, alpha-1A-, receptor
EPA	TOXCAST	NVS_GPCR_rAdra1 NonSelective	Adra1b	adrenergic, alpha-1B-, receptor
EPA	TOXCAST	NVS_GPCR_rAdra1 NonSelective	Adra1c	adrenergic, alpha-1C-, receptor
EPA	TOXCAST	NVS_GPCR_rAdra1 NonSelective	Adra1d	adrenergic, alpha-1D-, receptor
EPA	TOXCAST	NVS_GPCR_rAdra2 NonSelective	Adra2a	adrenergic, alpha-2A-, receptor
EPA	TOXCAST	NVS_GPCR_rAdra2 NonSelective	Adra2b	adrenergic, alpha-2B-, receptor
EPA	TOXCAST	NVS_GPCR_rAdra2 NonSelective	Adra2c	adrenergic, alpha-2C-, receptor
EPA	TOXCAST	NVS_GPCR_rAdrb NonSelective	Adrb1	adrenergic, beta-1-, receptor

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	NVS_GPCR_rAdrb NonSelective	Adrb2	adrenergic, beta-2-, receptor, surface
EPA	TOXCAST	NVS_GPCR_rAdrb NonSelective	Adrb3	adrenergic, beta-3-, receptor
EPA	TOXCAST	NVS_GPCR_rCRF	Crh	corticotropin releasing hormone
EPA	TOXCAST	NVS_GPCR_rGABAB	Gabbr2	gamma-aminobutyric acid (GABA) B receptor, 2
EPA	TOXCAST	NVS_GPCR_rGHB	GPR172A	G protein-coupled receptor 172A
EPA	TOXCAST	NVS_GPCR_rH3	Hrh3	histamine receptor H3
EPA	TOXCAST	NVS_GPCR_rmAdra2B	Adra2b	adrenergic, alpha-2B-, receptor
EPA	TOXCAST	NVS_GPCR_rmMGluR1	Grm1	glutamate receptor, metabotropic 1
EPA	TOXCAST	NVS_GPCR_rmMGluR5	Grm5	glutamate receptor, metabotropic 5
EPA	TOXCAST	NVS_GPCR_rNK1	Tac1	tachykinin, precursor 1
EPA	TOXCAST	NVS_GPCR_rNK3	Tacr3	tachykinin receptor 3
EPA	TOXCAST	NVS_GPCR_rNTS	NTSR1	neurotensin receptor 1 (high affinity)
EPA	TOXCAST	NVS_GPCR_rOpiate NonSelective	Oprl1	opiate receptor-like 1
EPA	TOXCAST	NVS_GPCR_rOpiate NonSelectiveNa	Oprl1	opiate receptor-like 1
EPA	TOXCAST	NVS_GPCR_rOXT	Oxtr	oxytocin receptor
EPA	TOXCAST	NVS_GPCR_rSST	Sstr1	somatostatin receptor 1
EPA	TOXCAST	NVS_GPCR_rTRH	Trhr	thyrotropin-releasing hormone receptor
EPA	TOXCAST	NVS_GPCR_rV1	Avpr1a	arginine vasopressin receptor 1A
EPA	TOXCAST	NVS_GPCR_rVIP Non_Selective	Vip	vasoactive intestinal peptide
EPA	TOXCAST	NVS_GPCR_rVIP Non_Selective	VIPR1	vasoactive intestinal peptide receptor 1
EPA	TOXCAST	NVS_GPCR_rVIP Non_Selective	VIPR2	vasoactive intestinal peptide receptor 2
EPA	TOXCAST	NVS_IC_bGABAAa1	GABRA1	gamma-aminobutyric acid (GABA) A receptor, alpha 1
EPA	TOXCAST	NVS_IC_bGABAAa5	GABRA5	gamma-aminobutyric acid (GABA) A receptor, alpha 5
EPA	TOXCAST	NVS_IC_bGABAAagonist	GABRA1	gamma-aminobutyric acid (GABA) A receptor, alpha 1
EPA	TOXCAST	NVS_IC_h5HT3	HTR3A	5-hydroxytryptamine (serotonin) receptor 3A
EPA	TOXCAST	NVS_IC_hKhERGCh	KCNH2	potassium voltage-gated channel, subfamily H (eag-related), member 2
EPA	TOXCAST	NVS_IC_hNNR_NBung Sens	CHRNA4	cholinergic receptor, nicotinic, alpha 4
EPA	TOXCAST	NVS_IC_rAMPA	Gria2	glutamate receptor, ionotropic, AMPA 2
EPA	TOXCAST	NVS_IC_rCaBTZCHL	Cacna1c	hypothetical protein LOC100131098; calcium channel, voltage-dependent, L type, alpha 1C subunit

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	NVS_IC_rCaChN	Cacna1b	calcium channel, voltage-dependent, N type, alpha 1B subunit
EPA	TOXCAST	NVS_IC_rCaDHPRCh_L	Cacna2d1	calcium channel, voltage-dependent, alpha 2/delta subunit 1
EPA	TOXCAST	NVS_IC_rGABAAa6	Gabra6	gamma-aminobutyric acid (GABA) A receptor, alpha 6
EPA	TOXCAST	NVS_IC_rGABANon Selective	Gabra1	gamma-aminobutyric acid (GABA) A receptor, alpha 1
EPA	TOXCAST	NVS_IC_rGluNMDA_MK80 1agonist	Grin1	glutamate receptor, ionotropic, N-methyl D-aspartate 1
EPA	TOXCAST	NVS_IC_rGluNMD Aagonist	Grin2a	glutamate receptor, ionotropic, N-methyl D-aspartate 2A
EPA	TOXCAST	NVS_IC_rGlyRStrySens	Glra1	glycine receptor, alpha 1
EPA	TOXCAST	NVS_IC_rKAR	Grik1	glutamate receptor, ionotropic, kainate 1
EPA	TOXCAST	NVS_IC_rKATPCh	Kcnj11	potassium inwardly-rectifying channel, subfamily J, member 11
EPA	TOXCAST	NVS_IC_rKCaCh	KCNN1	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 1
EPA	TOXCAST	NVS_IC_rKCaCh	KCNN2	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 2
EPA	TOXCAST	NVS_IC_rKCaCh	KCNN3	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3
EPA	TOXCAST	NVS_IC_rKCaCh	KCNN4	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4
EPA	TOXCAST	NVS_IC_rNaCh_site2	Scn1a	sodium channel, voltage-gated, type I, alpha subunit
EPA	TOXCAST	NVS_IC_rNNR_BungSens	Chrna7	CHRNA7 (cholinergic receptor, nicotinic, alpha 7, exons 5-10) and FAM7A (family with sequence similarity 7A, exons A-E) fusion; cholinergic receptor, nicotinic, alpha 7
EPA	TOXCAST	NVS_MP_hPBR	TSPO	translocator protein (18kDa)
EPA	TOXCAST	NVS_MP_rPBR	Tspo	translocator protein (18kDa)
EPA	TOXCAST	NVS_NR_bER	ESR1	estrogen receptor 1
EPA	TOXCAST	NVS_NR_bPR	PGR	progesterone receptor
EPA	TOXCAST	NVS_NR_hAR	AR	androgen receptor
EPA	TOXCAST	NVS_NR_hCAR	NR1I3	nuclear receptor subfamily 1, group I, member 3 (Constitutive Androstane Receptor)
EPA	TOXCAST	NVS_NR_hCAR_Agonist	NR1I3	nuclear receptor subfamily 1, group I, member 3 (Constitutive Androstane Receptor)
EPA	TOXCAST	NVS_NR_hER	ESR1	estrogen receptor 1
EPA	TOXCAST	NVS_NR_hFXR	NR1H4	nuclear receptor subfamily 1, group H, member 4 (Farnesoid X Receptor)
EPA	TOXCAST	NVS_NR_hGR	NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
EPA	TOXCAST	NVS_NR_hPPARa	PPARA	peroxisome proliferator-activated receptor alpha

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
EPA	TOXCAST	NVS_NR_hPPARg	PPARG	peroxisome proliferator-activated receptor gamma
EPA	TOXCAST	NVS_NR_hPR	PGR	progesterone receptor
EPA	TOXCAST	NVS_NR_hPXR	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
EPA	TOXCAST	NVS_NR_hRAR	RARA	retinoic acid receptor, alpha
EPA	TOXCAST	NVS_NR_hTRa	THRA	thyroid hormone receptor, alpha (erythroblastic leukemia viral (v-erb-a) oncogene homolog, avian)
EPA	TOXCAST	NVS_NR_rAR	AR	androgen receptor
EPA	TOXCAST	NVS_OR_gSIGMA NonSelective	SIGMAR1	sigma non-opioid intracellular receptor 1
EPA	TOXCAST	NVS_OR_hFKBP12	FKBP1A	FK506 binding protein 1A, 12kDa
EPA	TOXCAST	NVS_TR_gDAT	SLC6A3	solute carrier family 6 (neurotransmitter transporter, dopamine)
EPA	TOXCAST	NVS_TR_hAdoT	SLC29A1	solute carrier family 29 (nucleoside transporters), member 1
EPA	TOXCAST	NVS_TR_hDAT	SLC6A3	solute carrier family 6 (neurotransmitter transporter, dopamine), member 3
EPA	TOXCAST	NVS_TR_hNET	SLC6A2	solute carrier family 6 (neurotransmitter transporter, noradrenalin), member 2
EPA	TOXCAST	NVS_TR_hSERT	SLC6A4	solute carrier family 6 (neurotransmitter transporter, serotonin), member 4
EPA	TOXCAST	NVS_TR_rAdoT	Slc29a1	solute carrier family 29 (nucleoside transporters), member 1
EPA	TOXCAST	NVS_TR_rNET	Slc6a2	solute carrier family 6 (neurotransmitter transporter, noradrenalin), member 2
EPA	TOXCAST	NVS_TR_rSERT	Slc6a4	solute carrier family 6 (neurotransmitter transporter, serotonin), member 4
EPA	TOXCAST	NVS_TR_rVMAT2	Slc18a2	solute carrier family 18 (vesicular monoamine), member 2
EPA	TOXCAST	OxidativeStress	H2AFX	H2A histone family, member X
EPA	TOXCAST	StressKinase	JUN	jun oncogene
NCGC	EPA	PPAR-alpha	PPARA	peroxisome proliferator-activated receptor alpha
NCGC	EPA	PPARa protein interaction agonist mode	PPARA	peroxisome proliferator-activated receptor alpha
NCGC	EPA	PPARa protein interaction antagonist mode	PPARA	peroxisome proliferator-activated receptor alpha
NCGC	NTP	357	FOSB	FBJ murine osteosarcoma viral oncogene homolog B
NCGC	NTP	357	JUNB	jun B proto-oncogene
NCGC	NTP	357	JUND	jun D proto-oncogene
NCGC	NTP	357	JUN	jun oncogene
NCGC	NTP	357	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
NCGC	NTP	360	GBA2	glucosidase, beta (bile acid) 2

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
NCGC	NTP	360	GBA	glucosidase, beta; acid (includes glucosylceramidase)
NCGC	NTP	410	CYP1A2	cytochrome P450, family 1, subfamily A, polypeptide 2
NCGC	NTP	447	OGT	O-linked N-acetylglucosamine (GlcNAc) transferase (UDP-N-acetylglucosamine:polypeptide-N-acetylglucosaminyl transferase)
NCGC	NTP	530	MAPK10	mitogen-activated protein kinase 10
NCGC	NTP	595	HSP90AA1	heat shock protein 90kDa alpha (cytosolic), class A member 2; heat shock protein 90kDa alpha (cytosolic), class A member 1
NCGC	NTP	596	MAPT	microtubule-associated protein tau
NCGC	NTP	597	DNMT1	DNA (cytosine-5-)-methyltransferase 1
NCGC	NTP	662	CREB1	cAMP responsive element binding protein 1
NCGC	NTP	662	CREB3	cAMP responsive element binding protein 3
NCGC	NTP	875	BRCA1	breast cancer 1, early onset
NCGC	NTP	875	BACH1	BTB and CNC homology 1, basic leucine zipper transcription factor 1
NCGC	NTP	880	RGS12	regulator of G-protein signaling 12
NCGC	NTP	881	ALOX15	arachidonate 15-lipoxygenase
NCGC	NTP	883	CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9
NCGC	NTP	884	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4
NCGC	NTP	886	HSD17B10	hydroxysteroid (17-beta) dehydrogenase 10
NCGC	NTP	889	CASP7	caspase 7, apoptosis-related cysteine peptidase
NCGC	NTP	891	CYP2D6	cytochrome P450, family 2, subfamily D, polypeptide 6
NCGC	NTP	892	BRCA1	breast cancer 1, early onset
NCGC	NTP	892	BACH1	BTB and CNC homology 1, basic leucine zipper transcription factor 1
NCGC	NTP	893	HSD17B4	hydroxysteroid (17-beta) dehydrogenase 4
NCGC	NTP	894	HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)
NCGC	NTP	899	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19
NCGC	NTP	901	IMPA1	inositol(myo)-1(or 4)-monophosphatase 1
NCGC	NTP	902	TP53	tumor protein p53
NCGC	NTP	915	HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)
NCGC	NTP	924	TP53	tumor protein p53
NCGC	NTP	925	HBB	hemoglobin, beta
NCGC	NTP	926	TSHR	thyroid stimulating hormone receptor
NCGC	NTP	938	TSHR	thyroid stimulating hormone receptor
NCGC	NTP	995	MAPK1	mitogen-activated protein kinase 1

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
NCGC	NTP	1030	ALDH1A1	aldehyde dehydrogenase 1 family, member A1
NCGC	NTP	1452	ALOX12	arachidonate 12-lipoxygenase
NCGC	NTP	1458	SMN1	survival of motor neuron 1, telomeric; survival of motor neuron 2, centromeric
NCGC	NTP	1458	SMN2	survival of motor neuron 2, centromeric
NCGC	NTP	1461	NPSR1	neuropeptide S receptor 1
NCGC	NTP	1466	GAA	glucosidase, alpha; acid
NCGC	NTP	1467	GAA	glucosidase, alpha; acid
NCGC	NTP	1469	THRB	thyroid hormone receptor, beta (erythroblastic leukemia viral (v-erb-a) oncogene homolog 2, avian)
NCGC	NTP	1471	HTT	huntingtin
NCGC	NTP	1705	APEX1	APEX nuclease (multifunctional DNA repair enzyme) 1
NCGC	NTP	2120	HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)
NCGC	NTP	2546	RORC	RAR-related orphan receptor C
NCGC	NTP	2549	RECQL	RecQ protein-like (DNA helicase Q1-like)
NCGC	NTP	2551	RORC	RAR-related orphan receptor C
NCGC	NTP	5lo1911	ALOX5	arachidonate 5-lipoxygenase
NCGC	NTP	CRE (CHO)	CREB1	cAMP responsive element binding protein 1
NCGC	NTP	CRE (CHO)	CREB3	cAMP responsive element binding protein 3
NCGC	NTP	Endotoxin (agonist)	TLR4	toll-like receptor 4
NCGC	NTP	hERG	KCNH2	potassium voltage-gated channel, subfamily H (eag-related), member 2
NCGC	NTP	NFkB	NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
NCGC	NTP	P53-BLA	TP53	tumor protein p53
NCGC	NTP	P53-Caspase-3/7	TP53	tumor protein p53
NCGC	NTP	PPAR-gamma (CHO)	PPARG	peroxisome proliferator-activated receptor gamma
NCGC	NTP	PXR (DPX2)	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
NCGC	NTP	SF1	NR5A1	nuclear receptor subfamily 5, group A, member 1(steroidogenic factor 1)
NCGC	NTP-subset	AR protein interaction	AR	androgen receptor
NCGC	NTP/EPA	900	CASP1	caspase 1, apoptosis-related cysteine peptidase (interleukin 1, beta, convertase)
NCGC	NTP/EPA	AHR	AHR	aryl hydrocarbon receptor
NCGC	NTP/EPA	AR agonist	AR	androgen receptor

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
NCGC	NTP/EPA	AR antagonist	AR	androgen receptor
NCGC	NTP/EPA	ARE-BLA	AR	androgen receptor
NCGC	NTP/EPA	ARE-luciferase	AR	androgen receptor
NCGC	NTP/EPA	Caspase 8	CASP8	caspase 8, apoptosis-related cysteine peptidase
NCGC	NTP/EPA	Caspase 9	CASP9	caspase 9, apoptosis-related cysteine peptidase
NCGC	NTP/EPA	ER alpha agonist	ESR1	estrogen receptor 1
NCGC	NTP/EPA	ER alpha antagonist	ESR1	estrogen receptor 1
NCGC	NTP/EPA	ESRE	ESR1	estrogen receptor 1
NCGC	NTP/EPA	FXR agonist	NR1H4	nuclear receptor subfamily 1, group H, member 4 (Farnesoid X Receptor)
NCGC	NTP/EPA	FXR antagonist	NR1H4	nuclear receptor subfamily 1, group H, member 4 (Farnesoid X Receptor)
NCGC	NTP/EPA	GR agonist	NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
NCGC	NTP/EPA	GR antagonist	NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
NCGC	NTP/EPA	HSP-BLA	HSF3	heat shock factor 3
NCGC	NTP/EPA	HSP-BLA	HSF1	heat shock transcription factor 1
NCGC	NTP/EPA	HSP-BLA	HSF2	heat shock transcription factor 2
NCGC	NTP/EPA	HSP-BLA	HSF4	heat shock transcription factor 4
NCGC	NTP/EPA	HSP-BLA	HSF5	heat shock transcription factor family member 5
NCGC	NTP/EPA	HSP-luciferase	HSF3	heat shock factor 3
NCGC	NTP/EPA	HSP-luciferase	HSF1	heat shock transcription factor 1
NCGC	NTP/EPA	HSP-luciferase	HSF2	heat shock transcription factor 2
NCGC	NTP/EPA	HSP-luciferase	HSF4	heat shock transcription factor 4
NCGC	NTP/EPA	HSP-luciferase	HSF5	heat shock transcription factor family member 5
NCGC	NTP/EPA	LXR beta agonist	NR1H2	nuclear receptor subfamily 1, group H, member 2(Liver X receptor beta)
NCGC	NTP/EPA	LXR beta antagonist	NR1H2	nuclear receptor subfamily 1, group H, member 2(Liver X receptor beta)
NCGC	NTP/EPA	NFkB, agonist	NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
NCGC	NTP/EPA	NFkB, antagonist	NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
NCGC	NTP/EPA	PPAR-delta agonist	PPARD	peroxisome proliferator-activated receptor delta
NCGC	NTP/EPA	PPAR-delta antagonist	PPARD	peroxisome proliferator-activated receptor delta
NCGC	NTP/EPA	PPAR-gamma agonist	PPARG	peroxisome proliferator-activated receptor gamma

Assay Source	Library Tested	SOURCE_ NAME_AID	Name (gene symbol)-[homo sapiens]	Official Full Name
NCGC	NTP/EPA	PPAR-gamma antagonist	PPARG	peroxisome proliferator-activated receptor gamma
NCGC	NTP/EPA	PXR (human protein)	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
NCGC	NTP/EPA	PXR (Rat)	NR1I2	nuclear receptor subfamily 1, group I, member 2(Pregnane X Receptor)
NCGC	NTP/EPA	RXR agonist	RXRA	retinoid X receptor, alpha
NCGC	NTP/EPA	RXR antagonist	RXRA	retinoid X receptor, alpha
NCGC	NTP/EPA	TR-beta agonist	THRB	thyroid hormone receptor, beta (erythroblastic leukemia viral (v-erb-a) oncogene homolog 2, avian)
NCGC	NTP/EPA	TR-beta antagonist	THRB	thyroid hormone receptor, beta (erythroblastic leukemia viral (v-erb-a) oncogene homolog 2, avian)
NCGC	NTP/EPA	VDR	VDR	vitamin D (1,25- dihydroxyvitamin D3) receptor
NCGC	NTP/EPA subset	LXR beta	NR1H2	nuclear receptor subfamily 1, group H, member 2(Liver X receptor beta)
NCGC	NTP/EPA subset	LXRbeta protein interaction	NR1H2	nuclear receptor subfamily 1, group H, member 2(Liver X receptor beta)

Abbreviation: AID = assay identification number used on PubChem; gene annotation is by staff at the EPA, the NCGC, or NTP. Explaination of U.S. EPA ToxCast™ codes can be found at http://www.epa.gov/ncct/toxcast/assays.html

III.4 Tox21 Informatics Working Group

This page left intentionally blank

III.4 Tox21 Informatics Working Group

Keith R. Shockley, Ph.D.
NIEHS Biostatical Branch

Tox21 Informatics Working Group
Ruili Huang, Ph.D. (NHGRI NCGC)*
Mark Johnson, Ph.D. (U.S. Army)
Richard Judson, Ph.D. (U.S. EPA NCCT)*
David Reif, Ph.D. (U.S. EPA NCCT)
Keith Shockley, Ph.D. (NIEHS BB)*
Weida Tong, Ph.D. (U.S. FDA NCTR)*
** = Co-Chairs*

Biomolecular Screening Branch Scientists
Scott S. Auerbach, Ph.D., DABT
Windy A. Boyd, Ph.D.
Jonathan H. Freedman, Ph.D.
Fred M. Parham, Ph.D.

Associated NIEHS Scientists
Grace E. Kissling, Ph.D. (BB)
Changwon Lim, Ph.D. (BB)
Shyamal D. Peddada, Ph.D. (BB)

Collaborating Scientists
Patrick W. Crockett, Ph.D. (SRA International, Inc.)
Xiaohung Gu, Ph.D. (Kelly Scientific Services)
Deepak Mav, Ph.D. (SRA International, Inc.)
Ruchir R. Shah, Ph.D. (SRA International, Inc.)
Marjo V. Smith, Ph.D. (SRA International, Inc.)
Noel T. Southall, Ph.D. (NHGRI NCGC)
Dan Svoboda, Ph.D. (SRA International, Inc.)

III.4.1 Generation of Quantitative High Throughput Screening Data

III.4.1.1 Background

The goals of the Tox21 collaboration are to identify mechanisms of toxicity, prioritize chemicals for *in vivo* testing, and predict adverse responses to environmental chemicals in humans (Collins et al. 2008, Shukla et al. 2010). Past approaches to discover new medicinal compounds through low throughput animal and tissue models eventually gave way to HTS methods, which enables the simultaneous assessment of large numbers compounds. Generally, in drug discovery, HTS assays are conducted at a single test concentration with the goal of identifying compounds with strong pharmacological activity, an approach unsuitable for toxicological research. However, qHTS assays provide an opportunity to meet Tox21 objectives, holding the potential for wide chemical coverage and reduced cost of testing on

a per-substance basis. Moreover, the ability of a substance to induce a toxicological response is better understood by analyzing the response profile over a broad concentration range than by evaluating effects at one or a few concentrations.

The Tox21 collaboration began formally in 2008 with Phase I (Proof of Concept) consisting of qHTS studies conducted at the NCGC in 1536-well format and mid-throughput studies conducted in support of the U.S. EPA ToxCast™ program. In conjunction with Tox21 Phase I, the NTP and EPA have produced an extensive set of concentration response data on some 2800 substances screened at the NCGC in over 50 qHTS assays and on 320 substances tested across more than 500 *in vitro* and lower organism *in vivo* assays by various contract and government laboratories. Analyses of Phase I data indicate reproducible levels of compound behavior that match previously known toxicological responses across numerous assay conditions. Nevertheless, efforts devoted to the generation of qHTS data at the NCGC have led to an appreciation for analyzing the data in high-dimensional context and provided an opportunity for methodological development of alternative algorithms for making activity calls. Before attempting to make activity calls with qHTS data, it is necessary to assess data quality and account for important experimental factors (e.g., plate, row, and column effects). Typically, positive and negative activity calls are made by applying a suitable algorithm to normalized concentration-response data (see "Making Activity Calls" below), although normalization and activity testing can also be accomplished within the context of a mathematical model (e.g., Parham et al. 2009).

III.4.1.2 Tox21 Phase I efforts at the NCGC

The first stage of analysis of Tox21 data involved the examination of 1408 substances selected by the NTP (NTP-1408) and 1408 substances selected by the U.S. EPA (EPA-1408) in 1536 well plates. The plate design allows for the screening of a total of 1,408 compounds in a single 1,536 well plate; the remaining 128 wells are used for negative and positive controls. An additional 54 substances were later added by the U.S. EPA (EPA-54). Chemicals were tested for activity in >50 biochemical and cell-based assays (e.g., nuclear receptors [AR, ERα, FXR, GR, LXR, PPARα, PPARγ, TRβ, VDR] in agonist and antagonist mode; viability assays in 13 cell types; biochemical assays to measure interaction with different Cytochrome P-450 enzymes).

Data structures for qHTS assays vary between efforts at the NTP, U.S. EPA and NCGC. In general, the NTP-1408 compound library was run at 14 different concentration levels, with only one response data point produced per concentration. However, during optimization for 1536-well format, some assays were run in triplicate (Xia et al. 2008). The EPA-1408 (or the EPA-54) uses 1x, 3x, or 6x replication and might have been run at as few as 3 concentrations (with replication) or as many as 15 concentrations (no replication). Each experimental run follows a stacked order scheme in which each well in the first one or two and last two or one plates are DMSO (solvent) and the testing order proceeds from the smallest tested concentration (e.g., ~0.5 nM) to the largest (e.g., ~100 μM). While the concentration for each of the compounds tested remains the same in each plate, the first two columns in the plate contain dilutions of the positive control, effectively producing concentration-response curves for positive control compounds in every plate. A typical plate design is shown in **Figure III.4-1**, where Doxorubicin and Tamoxifen are positive controls for the assay and "Basal" refers to the DMSO negative control. Data are collected from the plates using a plate reader.

❖ One-channel data

Cell viability and, in some cases, gene reporter assays operate through the measurement of a single readout (channel). For example, cell viability is measured using a luciferase-coupled adenosine triphosphate (ATP) quantitation assay (Cell-Titer-Glo®, Promega, Madison, WI), where the luminescent

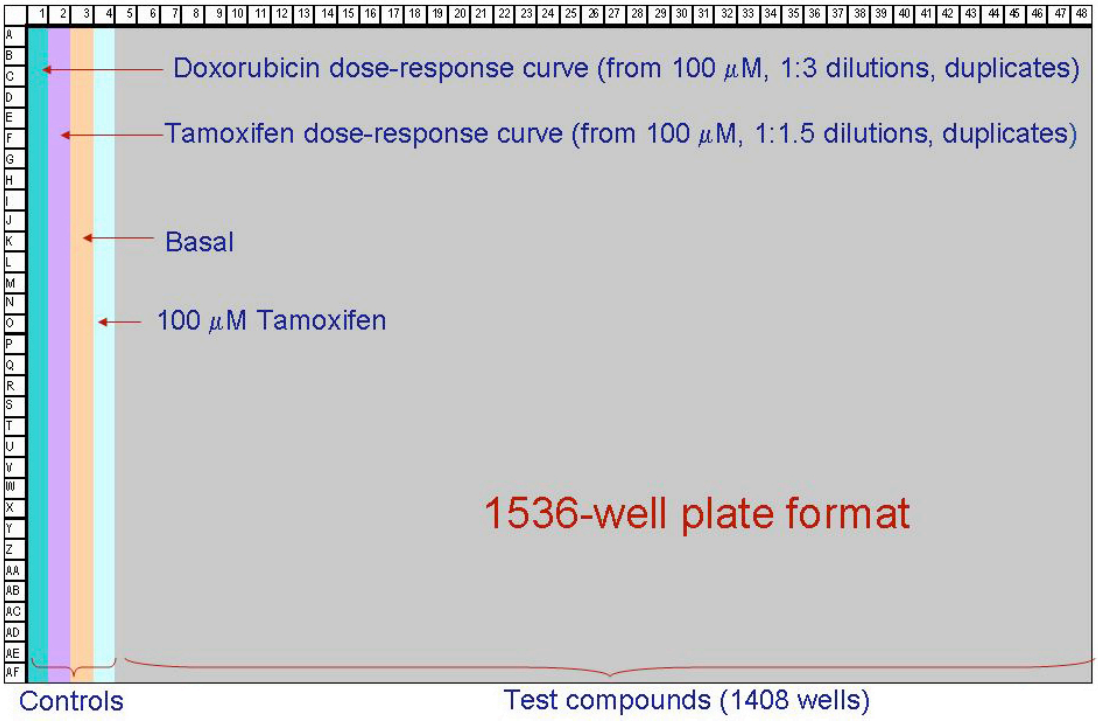


Figure III.4-1 Example assay test plate design for cytotoxicity assays

signal is proportional to the amount of ATP present and, correspondingly, to the number of metabolically active (or viable) cells.

Some gene reporter assays also may be read through a one-channel readout (e.g., Acumen®) in which the signal corresponds to the reporter signal. In both cases, the measured signal intensity values are proportional to the experimental response being studied.

❖ **Two-channel data**

Signal intensity data from some reporter gene assays may be obtained from the Förster resonance energy transfer (FRET)-based beta-lactamase assay system, which collects two emission measurements after excitation at 405 nm: green and blue. Green fluorescence emission at 530 nm (or channel 1) is produced when the FRET moiety is intact (i.e., no cleavage of FRET moiety by beta lactamase in cells). Therefore, emission at 530 nm is related to cell number. When the FRET moiety is cleaved by beta-lactamase under the control of the tested gene promoter, excitation results in blue fluorescence emission at 460 nm (channel 2). Blue emission indicates activity elicited by the interaction of a tested chemical and gene promoter system. A ratio channel is also calculated by the instrument as the ratio of signal from channel 2 divided by signal from channel 1.

❖ **Evaluating data quality**

Assay performance can be evaluated through a number of different statistical parameters, including signal-to-noise ratio, signal-to-background ratio, coefficient of variation, Z-factor and Z'-factor (Zhang et al. 1999, Inglese et al. 2007). However, both the signal-to-noise ratio and the signal-to-background ratio fail to assess the variability in sample and control measurements and the signal dynamic range. In

contrast, the Z-factor introduced by Zhang et al. (1999) can assess assay quality for a particular qHTS data set by accounting for assay signal dynamic range and variability with the negative control (or background) measurements. The Z-factor is defined as the calculated separation band (distance between 3 standard deviations from the mean of signal and 3 standard deviations of negative control signal) divided by the calculated assay dynamic range (distance between mean of the signal and mean of the control measurements). The Z'-factor is calculated using only control data, where the signal measurements are replaced with positive control measurements. Z-factor scores greater than 0.5 reflect excellent assays.

Intra-experimental assay reproducibility based on 55 compounds present in duplicate in each NTP 1408 assay plate was evaluated in 13 cell types using a cytotoxicity assay based on the measurement of ATP levels with Z'-factors ranging from 0.44 to 0.91 (Huang et al. 2008). In that study, the regression correlation between calculated IC_{50}[1] values for these replicates was found to be statistically significant ($R^2 = 0.71$, $p < 0.001$). Inter-experimental reproducibility was evaluated in the same study by comparing triplicate runs of 1,408 compounds on HepG2 cells. Of the 1,353 unique compounds in the study, 90% of the compounds had the same activity classification outcome in all three runs and an average $R^2 = 0.74$, $p < 0.001$). The regression correlations between IC_{50} values in pairs of runs was good with an average $R^2 = 0.74$ ($p < 0.001$) calculated from more than 1000 data points.

Certain compounds may auto fluoresce at the measured wavelength. In such cases, fluorescence signals are concentration dependent, regardless of activity of the compound. One way to perform a preliminary screen for such compounds is to test for a concentration-response in multiple assays at a given wavelength. For instance, fifteen compounds exhibited strong activity in at least 5 out of the 10 nuclear receptor agonist assays measured in the 460 nm channel (Channel 2), which gave rise to the suspicion that they are likely blue fluorescent. Some of these compounds are also fluorescent in the 530 nm channel (Channel 1), which may indicate green fluorescence. The auto fluorescence potential of 10 out of the 15 compounds was confirmed by spectra profiling experiments. The remaining 5 compounds were negative in the spectra profiling assay, making the call of auto fluorescence activity suspicious, but less conclusive. It is important to consider that an auto fluorescent compound may still be active in a tested assay (e.g., a real inducer of nuclear receptor activity). Testing compounds that are suspected to be auto fluorescent can be accomplished by testing the compound in non-fluorescence based orthogonal assays or attempting to detect the response at wavelengths outside of the auto fluorescence range of the compound being screened.

❖ Data normalization and outlier assessment

At the NCGC, raw plate reads for each titration point are first normalized relative to the positive control compound (agonist mode: 100%; antagonist mode: 0%) and DMSO-only wells (agonist mode: 0%; antagonist mode: -100%) as follows:

$$\text{Percent Activity} = \frac{V_{\text{compound}} - V_{\text{DMSO}}}{V_{\text{positive}} - V_{\text{DMSO}}} \times 100 \qquad \text{(Eqn. 1)}$$

where $V_{compound}$ denotes the compound well values, $V_{positive}$ denotes the median value of the positive control wells, and V_{DMSO} denotes the median values of the DMSO-only wells. These values are then corrected for row, column, and plate effects by applying a NCGC in-house pattern correction algorithm

[1] The inhibition concentration of a compound that induces a half-maximal decrease in a response

based on linear interpolation. Normalized response signals therefore can be regarded as the percent of response generated by the positive control.

Detecting outliers in nonlinear response data with limited replication is challenging. Current approaches define probable outliers based on the fit to a Hill Equation (Hill 1910). In general, we use a "leave-one-out" approach to assess whether the fit to the Hill Equation is improved by leaving out a particular data point. If the fit is improved, then the data point is marked as an outlier. This process is repeated for each data point at each tested concentration level, until all data points have been evaluated. The curves are then fit to the Hill Equation without the outlier data point(s). All further statistics are calculated without the data points in NTP efforts, or with the outlier data points in calculations performed at the NCGC (see below). The NCGC Assay Guidance Manual contains more extensive documents of assay validation and normalization procedures (http://www.ncgc.nih.gov/guidance/manual_toc.html).

Establishing detection limits provides a range within which there is a reasonable confidence in the meaning and reliability of a signal. In extreme cases, compounds may show spurious concentration-response behavior below detection thresholds (e.g., due to machine drift). In most cases, the detection limit is set as 3 standard deviations above or below the normalized signal values in negative control (DMSO-only) plates. In some cases, data from the lowest concentration plate (usually ~0.5 nM of test compounds) is combined with the negative control samples to calculate a detection limit. A detection limit of 25-30% of the positive control is a typical detection threshold for detecting reliable response signals in an assay.

III.4.1.3 Planned activities in Tox21

Phase II of Tox21 will utilize a >10,000 compound library for qHTS studies conducted at the NCGC. The plate design for Phase II will still allocate a total of 1,408 wells for test compounds in each 1,536 well plate, using the remaining 128 wells for controls as done previously. And, each of the 1,408 tested compounds will be present at the same concentration on the same plate, as was conducted during Phase I at the NCGC. However, in Phase II, intra-array replication will be assessed by including 88 duplicate compounds in each and every plate. Most importantly, unlike the studies in Phase I, each experimental run will be performed with each compound tested in at least triplicate concentration response curves, with plate orientation varying between each complete set of concentrations in order to better assess and account for well-to-well variability.

III.4.2 Making Activity Calls

The normalized and processed data set emerging from qHTS studies at the NCGC is very large; any assessment of the data that relies exclusively on manual inspection of individual curves is laborious, subjective, and prone to human error. Indeed, it is not "possible" for the human eye to reliably discriminate calls based on small (but statistically significant) trends/differences. A number of heuristic approaches and statistical models have been developed to address a variety of qHTS data structures derived from different types of studies (e.g., Inglese et al. 2005, Ritz and Streibig 2005, Parham et al. 2009). The determination of which approach is most appropriate may depend on *a priori* knowledge of the assay in question, the purpose of the study, or the structure of the data. Within Tox21 efforts, the choice of test method is usually determined by the intended use of outcomes; the methods described here are still being evaluated and revised as new questions arise. The detection of false positives can be minimized by restricting prioritization to chemicals with more robust concentration-response profiles.

Activity call algorithms used in Tox21 are most often based on nonlinear modeling of concentration-response relationships that follow a four-parameter Hill Equation (Hill 1910). The general form of the Hill Equation for agonist assays is presented below,

$$\text{response} = E_{init} + (E_{max} - E_{init})\frac{conc^n}{(EC_{50})^n + conc^n} \quad \text{(Eqn. 2)}$$

where the response is modeled using the four parameters E_{init} (the lowest activity or lower asymptote of the sigmoidal curve), E_{max} (the maximal activity or the upper asymptote of the curve), *conc* (the compound concentration), EC_{50} (the concentration exhibiting 50% of the maximal response), and *n* (the Hill coefficient that affects the shape of the curve). For antagonist and cytotoxicity assays, Eqn. 2 can be used where AC_{50} is replaced with IC_{50}, the concentration exhibiting 50% of the maximal inhibitory response. In agonist assays, the response is expected to increase with increasing concentration, while in antagonist or cytotoxicity assays, it is expected that the response will decrease with increasing concentration. The three-parameter Hill Equation is based on Eqn. 2, with $E_{init} = 0$.

III.4.2.1 Mathematical modeling of replicated one-channel data (NTP/NCGC)

A mathematical approach was developed to normalize qHTS data by removing bias due to plate location and to test for concentration-response relationships in one-channel cytotoxicity data generated within the cooperative HTS screening effort between NTP and NCGC (Parham et al. 2009). The approach tests for significance of response and quality of fit. It also estimates the functional parameters of the concentration-response curve (using the 3 parameter Hill Equation), and provides a method for categorizing compounds into different activity classes. Rather than adjusting for row and column effects within a plate prior to the analysis, the model used here adjusts for these effects as part of the formal analysis. The method was applied to cytotoxicity data from 1353 unique compounds in 9 human and 4 rodent cell types with 14 compound concentrations ranging from approximately 0.5 nM to 92 μM and in HepG2 cells, where the compound library was screened in three independent runs. This algorithm was able to fit the data well and remove plate location effects. The authors applied their approach to the HepG2 data to examine how reproducible the data are within runs based on duplicated substances and between runs based on HepG2 replicated experiments (Parham et al. 2009). The normalized data were highly correlated between duplicates or between triplicates at high concentration levels, although the concentration-response parameters were less highly correlated. The study also found that concentration-response curves with stronger responses tended to be more significantly correlated than curves with weaker responses.

III.4.2.2 Decision Tree Algorithm for multiple concentration data (NTP)

Pharmaceutical applications of qHTS generally seek to minimize false positives in activity assessment and often rely on heuristic algorithms to make activity calls. In contrast, multiple questions must be asked to determine the toxicological relevance of a chemical and statistical testing should be used to minimize false negatives in this setting. For instance, consider the questions below that are often asked with NTP investigations:

- Is there a robust concentration-response within the tested concentration range?
- Is there a concentration-response below the minimum tested concentration?
- Is there a weak concentration-response within tested concentration range?

In addition to these principal questions, sub-questions may also be of interest, including:

- Can the concentration-response be explained by cytotoxicity?
- Is the concentration-response correlated with cytotoxicity?
- What is the shape of the concentration-response curve?
- What is the lowest effective concentration (LEC) level?
- Is the concentration-response due to auto fluorescence?
- Is the concentration-response slope biologically plausible?

To compare activity calls between assay conditions or chemicals, it is also desirable to have an activity call algorithm that can accommodate different levels of replication and data types (e.g., one-channel or two-channel data, agonist and antagonist assays). To assess compounds in qHTS assays for toxicological research in the NTP, we developed a multiple-stage decision tree statistical model (see **Figure III.4-2**). This approach places chemicals into different bins describing the activity and statistical confidence of the response. In **Figure III.4-2**, active compounds are shown in red and inactive compounds are in green.

In this approach, data are fit to a four-parameter Hill equation and an overall F-test comparing the best fit to the Hill equation and a horizontal no response line is calculated for each chemical. Substances with a robust concentration-response are identified in the first stage (see TEST1 in **Table III.4-1**). In the second stage, compounds not detected as active in the first stage are evaluated for a maximal response at the lowest concentration by comparing the distribution of measured responses to a control value. Chemicals with a weak concentration-response are identified in the third stage, and the final stage separates substances exhibiting a cytotoxic response at the lowest concentration from inactive compounds. Significance levels for the tests conducted in the Decision Tree algorithm are typically set to $\alpha = 0.05$. When large compound libraries are interrogated, a false discovery rate less than 5% is used for each test. The direction of the response is indicated within TEST calls (e.g., "ACTIVE*[1]" would be active in TEST 1 and the response increases with increasing concentration). Known and suspected autofluorescent compounds are classified as "INACTIVE*", unless the purpose of the analysis is to find candidate autofluors.

A program called *SpreadSheetGen* has been designed by the NTP for the R programming language and software environment (http://www.R-project.org). This program requires R/drc (Ritz and Streibig, 2005) to fit the data to the Hill equation and can be run on Microsoft® Windows or Linux server platforms. *R/SpreadSheetGen* analyzes qHTS data by reading a series of text files containing concentrations and normalized responses and generates spreadsheets with calculated p-values, Hill equation parameters, and other results. The resulting spreadsheets can then be sorted according to the NTP Decision Tree strategy. A separate program called, "NTP qHTS Java Plot Generator" was developed by the NTP to produce summary plot files in JPEG format based on results generated by *SpreadSheetGen*. Each JPEG file contains one or more plots showing the concentration-response values and the fitted hill curve (when appropriate). The list of chemicals to be plotted is specified by a tab-delimited text file supplied by the user.

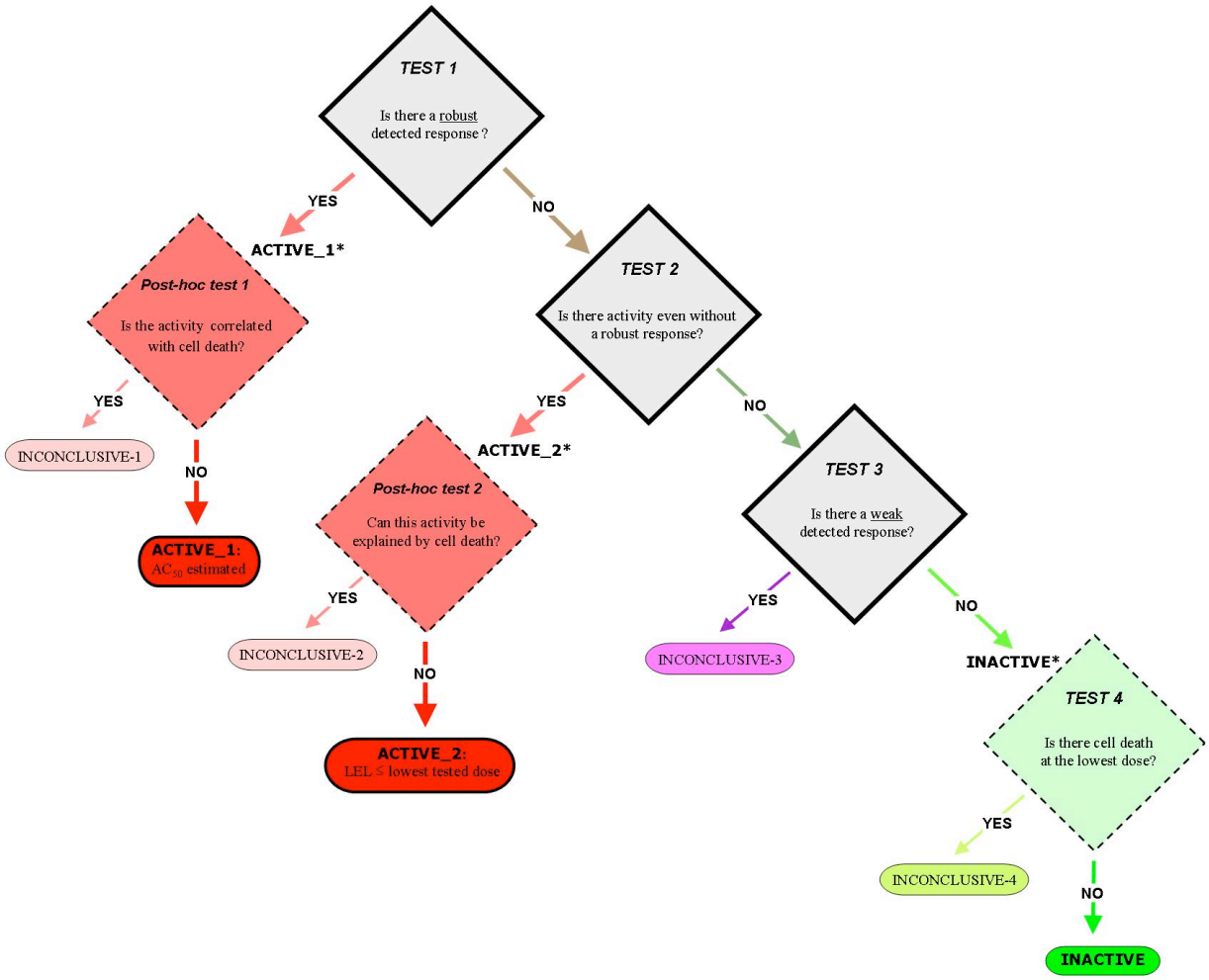


Figure III.4-2 Decision Tree Algorithm

❖ *Curve fit visualization tool to investigate weak activity calls (NTP)*

The curve calling algorithms discussed previously may yield inconclusive calls, such as compounds classified as "INCONCLUSIVE[3]" in the NTP Decision Tree algorithm (**Figure III.4-2**) or curve class "3" in the NCGC approach (see Tables III.3-2 and III.3-3). In these cases, visual inspection of the relevant concentration-response curves may help to refine or clarify the activity calls. A graphical user interface application enables interactive viewing of concentration-response data, together with curve parameters derived from fits to the Hill Equation. Users must log in to the application to identify themselves, and then they are able to browse through selected datasets. If desired, a user may choose to designate an inconclusive dataset as positive or negative. The user specified calls are recorded along with the caller's identity. After all reviewers have made calls on a relevant experiment, the experiment is flagged with

Table III.4- 1 Tests within Decision Tree Algorithm

Test	H_0	Ha
TEST1 *(F-test: WLS[a] and OLS[b])*	$\beta_{hill+} = 0$ $\beta_{hill-} = 0$	$\beta_{hill+} > 0$ $\beta_{hill-} < 0$
PH1 *(Spearman correlation)*	$\rho_{resp+/tox} = 0$ $\rho_{resp-/tox} = 0$	$\rho_{resp+/tox} > 0$ $\rho_{resp-/tox} < 0$
TEST2 *(weighted t-test)*	μ_{resp+} = noise μ_{resp-} = -noise	μ_{resp+} > noise μ_{resp-} < -noise
PH2 *(classical t-test)*	$\mu_{resp+} = -\mu_{tox}$ $\mu_{resp-} = \mu_{tox}$	$\mu_{resp+} > -\mu_{tox}$ $\mu_{resp-} < \mu_{tox}$
TEST3 *(F-test: WLS[a] or OLS[b])*	$\beta_{hill+} = 0$ $\beta_{hill-} = 0$	$\beta_{hill+} > 0$ $\beta_{hill-} < 0$
TEST4 *(classical t-test)*	μ_{tox+} = noise μ_{tox-} = -noise	μ_{tox+} > noise μ_{tox-} < -noise

**Tests for agonist (+) or antagonist [or toxic] (-) activity*
[a]*WLS refers to weighted least squares*
[b]*OLS refers to ordinary least squares*

designations “callers concur”, or “callers do not concur”. In cases where callers do not concur, disagreement is resolved by discussion. In this way, ambiguous calls are reviewed until all flags read “callers concur”.

III.4.2.3 Assay-specific analysis approaches (U.S. EPA)

There are many different HTS data structures (e.g., different numbers of tested concentrations, different levels of replication per data point) within U.S. EPA’s ToxCast™ database (http://www.epa.gov/ncct/toxcast). When concentration-response data can be reliably described by the Hill equation (e.g., when a lower and an upper boundary of the curve fit to the Hill equation can be confidently modeled), half maximal concentration values (EC_{50} or IC_{50}) are calculated from curve fits to the three- or four-parameter Hill equation. In assay sets in which no upper asymptote (agonist assays) or lower asymptote (antagonist or cytotoxicity) can be defined, a LEC is calculated. The LEC is defined as the lowest concentration at which there is a statistically significant difference from the concurrent negative control. LEC values are typically smaller than calculated EC_{50} or IC_{50} values, because an LEC can usually be determined even if 50% activity is never reached within the concentrations tested. Results are placed into internal databases where they are integrated with data linking the qHTS data with genes, pathways, and other endpoints. The results are filtered based on cytotoxicity, and parameters such as concentration for half maximal activity ($AC_{50} = EC_{50}$ or IC_{50}), maximum efficacy (E_{max}), and minimal response (E_{init}) are used to make activity calls based on in-house algorithms for specific assay types and platforms. A more extensive description of curve-fitting procedures for various assay types and analysis settings can be found in Judson et al. (2010) and in the supplemental information accompanying that article.

III.4.2.4 Curve class approaches for multiple concentration Data (NCGC)

Historically, the NCGC used curve classes, heuristic measures of data confidence, to classify concentration-responses on the basis of efficacy, the number of data points observed above background activity, and the quality of fit. Curve class designations are shown in **Table III.4-2**.

Table III.4-2 Curve class designations (modified from Inglese et al. 2006)

Curve Class	Description	Efficacy	r^2	Asymptotes	Inflection
1.1	Complete curve	>80%	≥0.9	2	Yes
1.2	Complete curve	≤80%	≥0.9	2	Yes
1.3	Complete curve	>80%	<0.9	2	Yes
1.4	Complete curve	≤80%	<0.9	2	Yes
2.1	Incomplete curve	>80%	≥0.9	1	Yes
2.2	Incomplete curve	≤80%	<0.9	1	Yes
2.3	Incomplete curve	>80%	<0.9	1	Yes
2.4	Incomplete curve	≤80%	≥0.9	1	Yes
3	Single point activity	>3SD	NA	1	No
4	Inactive	≤3SD	NA	0	No

The NCGC curve class method has recently been revised to better suit toxicological research. The amended NCGC method now incorporates p-values (derived from an F-test that measures the curve fit quality) into the curve classification system. Also, a new curve class (class 5) has been added to the categorization in order to describe chemicals with activity at the lowest tested concentration. Among the activators/inhibitors, compounds with class 1.1, 1.2, 2.1 or 2.2 curves ($p<0.05$) and >60% efficacy in the ratio readout were further defined as active activators or inhibitors. Compounds that were class 4 in both the ratio and 460 nm readouts were defined as inactive and compounds with activity at the lowest concentration are classified as inconclusive. The new curve class system is presented in **Table III.4-3**.

III.4.2.5 Future directions in activity assessment

There is room to improve current activity call methods and develop new algorithms that are more appropriate to novel or more complex questions. For instance, comparing the activities of two chemical profiles has so far relied on pattern recognition algorithms based on AC_{50} values derived from the Hill equation (e.g., a heatmap containing calculated potencies for each chemical and assay combination). While this procedure is useful for finding global correlations in activity patterns, differences between curve fit parameters derived from the Hill equation have not been compared with formal statistical testing in this scenario and formal comparisons may be more appropriate when considering a small number of secondary hypotheses. Efforts to develop these statistical procedures to answer these kinds of questions are currently underway within the NTP. Recently, the U.S. EPA has begun a curve-fitting effort that uses Bayesian techniques to assign credibility intervals to all calculated parameters. Only parameters with narrow credibility intervals would be considered sufficiently robust to use as a basis of activity assessment or downstream analyses.

Table III.4-3 Revised curve class designations

Curve Class	Description	Efficacy	*p*-value*	Asymptotes	Inflection
1.1	Complete curve	>6SD[†]	<0.05	2	Yes
1.2	Complete curve	≤6SD; >3SD	<0.05	2	Yes
1.3	Complete curve	>6SD	≥0.05	2	Yes
1.4	Complete curve	≤6SD; >3SD	≥0.05	2	Yes
2.1	Incomplete curve	>6SD	<0.05	1	Yes
2.2	Incomplete curve	≤6SD; >3SD	<0.05	1	Yes
2.3	Incomplete curve	>6SD	≥0.05	1	Yes
2.4	Incomplete curve	≤6SD; >3SD	≥0.05	1	Yes
3	Single point activity	>3SD	NA	1	No
4	Inactive	≤3SD	≥0.05	0	No
5[‡]	Inconclusive	NA	NA	NA	NA

Abbreviations: NA = not applicable; SD = standard deviation

* *p*-value is derived from an overall F-test that measures the quality of curve fit.

[†] SD of sample activities at the lowest tested concentration and values of the DMSO control wells.

[‡] Class 5 is a special class for samples with activity at the lowest tested concentration (>6SD or >3SD and the difference between the maximal change in activity observed in the tested concentration range and the negative controls is <3SD.

III.4.3 Chemical Prioritization for Toxicity Testing

Assessing health risks of an environmental chemical generally proceeds through *in vivo* bioassays that take, at a minimum, several years to complete, costing millions of dollars. However, there are an estimated 30,000 unique chemicals in wide commercial use (Muir and Howard 2006, Judson et al. 2008), and most of these chemicals have not yet been tested. Accordingly, the central goal of chemical prioritization efforts is to use information derived from chemical-activity data in HTS to prioritize chemicals for toxicity testing or further studies. This prioritization can be based on activity patterns generated from a toxicologically relevant assay or chemical profiles related to common mechanistic targets that are derived from many assays. To identify mechanistic targets, it is necessary to screen a diversity of compounds, representing a variety of structures and properties, and determine which compound sets have similar activity patterns. Mechanisms might be identified using qHTS assays to screen for active compound sets before proceeding to traditional *in vivo* toxicity assays or other studies by using staged evaluations (Huang et al. 2008). In addition, prediction models can be developed based on qHTS to complement staged evaluations by confirming chemical signatures (i.e., chemical activity patterns) or understanding toxicological mechanisms. Different entities within Tox21 have utilized different approaches for prioritization, and many of the methods described below are relatively new and are still being evaluated.

III.4.3.1 Using in vitro activity patterns for staged evaluations

- **Rank-ordering (NTP)**

The NTP uses multiple-concentration qHTS data sets to investigate concentration-response relationships for qHTS assays. Highly active compounds can be determined based on the curve fit to the four-parameter Hill Equation and a p-value for significance of the fits can be determined for each chemical (see “Mathematical Modeling of Replicated One-Channel Data” or “ACTIVE*[1]” or “ACTIVE*[-1]” activities in “Decision Tree Algorithm for Multiple Concentration Data” above). Significant responses at

the lowest tested concentration are detected as "ACTIVE*[2]" or "ACTIVE*[-2]" compounds in the decision tree approach (see **Figure III.4-2**). The set of active chemicals can be ranked according to significance criteria (false discovery rate, p-value, or F-statistic) or by a curve fit parameter (e.g., the potency or maximal response derived from a fit to the Hill equation). It also may be useful to rank the compounds according to a calculated parameter derived from the potency (e.g., AC_{50} or IC_{50}) and maximal response (e.g., E_{max} or E_{init}), such as:

$$\text{RankStat}_{\text{activator}} = \frac{|E_{max}|}{AC_{50}}, \qquad \text{RankStat}_{\text{inhibitor}} = \frac{|E_{init}|}{IC_{50}} \qquad \text{(Eqn. 3)}$$

which represents the ranking statistics for activator (increasing activity with increasing concentration) and inhibitor (decreasing activity with increasing concentration) assays, respectively. Chemicals with large maximal responses ($|E_{max}|$ or $|E_{init}|$ is large) or with very small potencies (AC_{50} or IC_{50} is very small) will result in larger values of the ranking statistics than other chemicals.

❖ Comparing chemical profiles or assays (NTP)

In some cases, toxicological mechanism can be investigated by comparing the response profiles of two chemical profiles within a single assay or between two assays. For instance, consider the two chemical profiles shown below generated from data in an AR agonist assay (**Figure III.4-3**). The red curves represent the normalized response signals from the assay, and the green curves represent a measure of cytotoxicity. The chemicals were analyzed using the NTP Decision Tree Algorithm described earlier. The Spearman coefficient describing the degree of correlation between the red and green curves is shown on each plot. In **Figure III.4-3A**, Androstenedione shows a significant agonist concentration-response in the AR assay, independent of cytotoxicity, while in **Figure III.4-3B**, the concentration-response profile of Croton oil in the AR agonist assay is inversely correlated with cytotoxic response ($p < 0.0001$). Both chemicals are designated as actives, but the concentration-response profile of Croton oil is associated with a cytotoxic response, which might be of interest.

Similarly, mechanistic understanding of chemical effect may depend on dissimilar responses in two different assays. For instance, consider the concentration-response profiles generated in a cytotoxicity assay conducted in a wild type cell line (shown in green) versus an assay conducted in the same cell type but with a particular gene knock-out (shown in red). In **Figure III.4-4A**, 4-(Chloroacetyl)acetanilide shows a strong cytotoxic response in both cell systems. In contrast, in **Figure III.4-4B**, the knockout cells show strong cytotoxicity, but the wild type cells remain viable over the entire range of tested concentration. The differences in concentration-response between the two cell lines lead to the hypothesis that the target gene plays an important role in response to Acetochlor.

High-dimensional data analysis methods may also be used to define activity categories for compound sets. For instance, it may be advantageous to compare response profile a chemical of interest, such as Bisphenol A (BPA), to response profiles of other tested chemicals with unknown activities. To begin such a comparison, we use the NTP Decision Tree Algorithm to find chemicals with a positive activity response (designated "ACTIVE*[1]" or "ACTIVE*[2]"). An AC_{50} value can be calculated for each concentration-response curve associated with an active chemical. A 1408 x 9 matrix containing AC_{50} values was calculated from Hill equation fits to data from 9 nuclear receptor agonist assays (AR, ERα, FXR, GR, PPARα, PPARγ, RXR, TRβ, and VDR). This matrix was then filtered to 95 substances that were

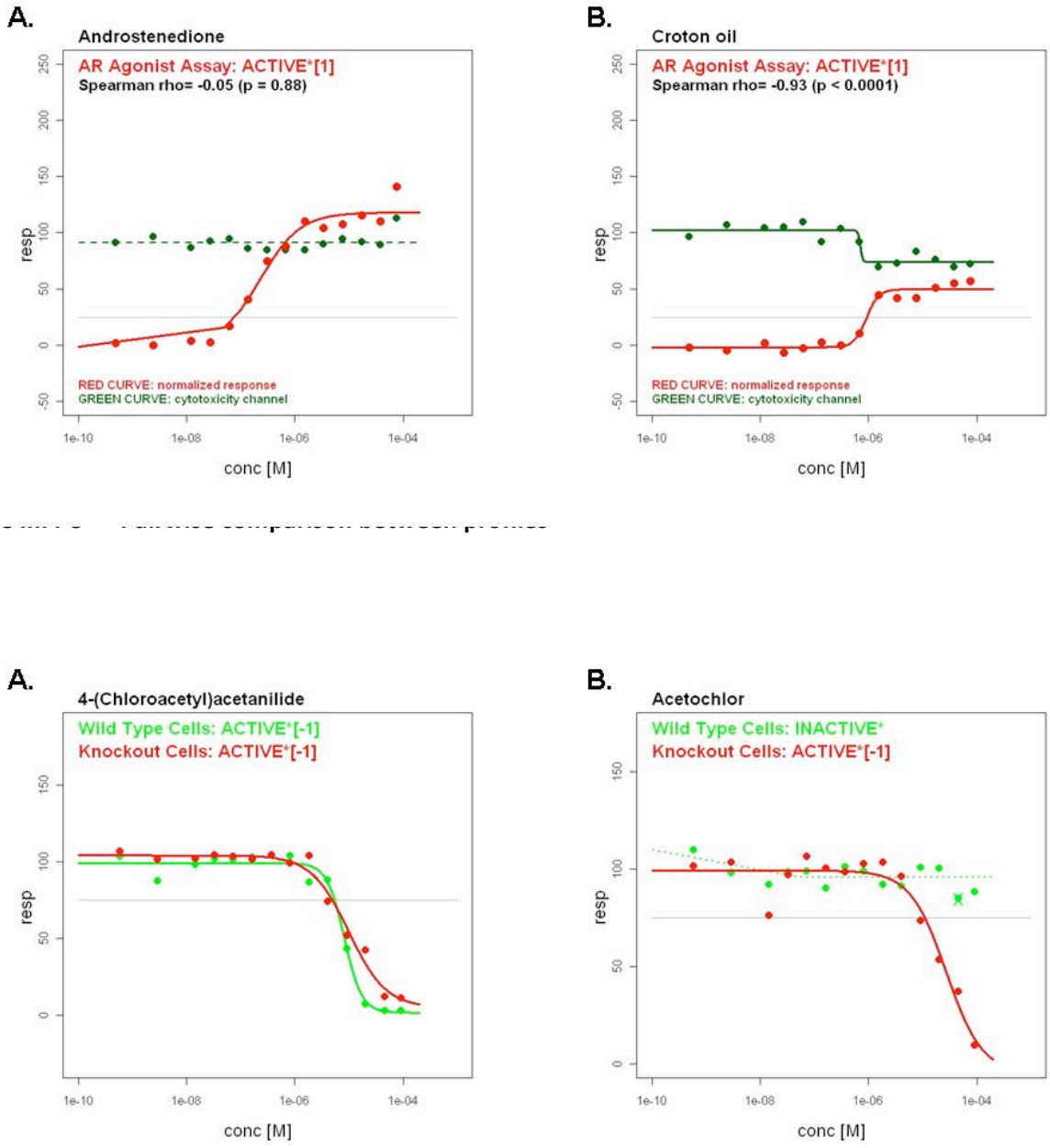


Figure III.4-4 Pairwise comparison between assays

active in at least one assay. These chemicals were subsequently grouped into 6 bins according to their AC_{50}: [1] NOT ACTIVE, [2] >50 µM, [3] 10-50 µM, [4] 1-10 µM, [5] <1 µM, and [6] <<1 µM. The rows and columns of the resulting 95 x 9 matrix were subjected to hierarchical cluster analysis and the resulting heat map is shown in **Figure III.4-5**.

The ERα assay produced the most active compounds (39), while PPARγ produced the next highest number of actives (25). As shown in **Figure III.4-5**, some chemicals are active in only one assay while other chemicals are active across multiple assays. A number of different chemicals shared similar activity

profiles across the 9 assays. Chemicals with the same activity profiles can be grouped into compound sets for downstream analyses. Compound sets may be used for prediction modeling in order to find substructures within the list that may be responsible for a particular activity. Alternatively, compound sets may be used to prioritize chemicals with unknown activity based on the activity of a compound with a known adverse response. The profiles corresponding to two substances representing BPA and three chemicals with very high potencies (i.e., very low AC_{50} values) are highlighted in **Figure III.4-5**. Within this collection of qHTS screens, only ERα activity was noticeably perturbed by BPA, but other chemicals showed this same activity pattern.

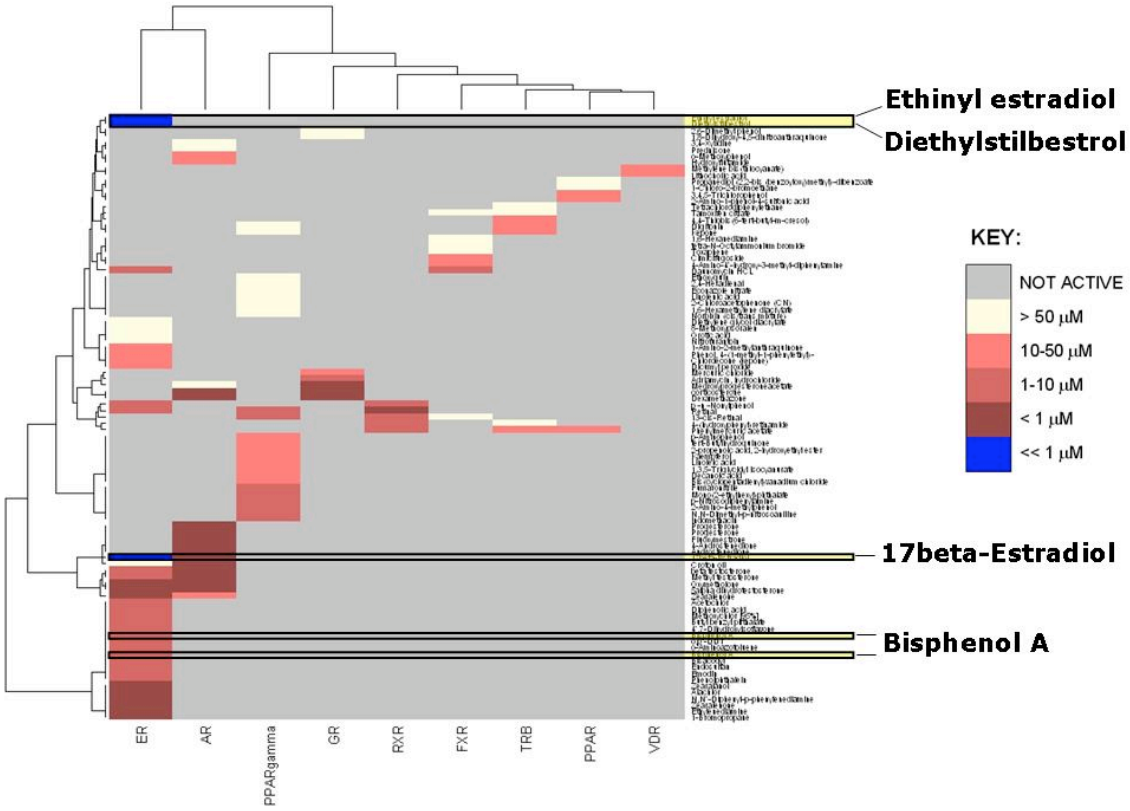


Figure III.4-5 Heatmap for assessing multiple profiles

❖ WormTox modeling (NTP)

The nematode *C. elegans* is an alternative animal model considered for a role in a multi-tier toxicant screen. The 320 ToxCast™ Phase I chemicals were tested using a 7 dose growth and development assay in these nematodes (see Section 4.1-1: *Caenorhabditis elegans* "Worm Tox" Screening Facility for a more extensive discussion). The data generated on *C. elegans* provides an opportunity for cross species comparisons (*C. elegans* versus rodent developmental and cancer results on the same chemicals). To compare the results generated by the nematode studies to the other species, comparable measures of toxicity were chosen from each species. After looking at several measures of activity, the most sensitive measure of toxicity for the nematode assay was an activity score that incorporated both the steepness

of the dose-response curve and information on the average size of nematodes at each dose measured as percent of control values. The Kendall tau statistic was used to compare these activity scores and lowest effect level (LEL) values found in ToxRefDB. Only chemicals with an LEL listed for at least one of 6 rodent endpoints were included in the analysis. In spite of the discrete nature of the measures compared, these results show significant consistency between rodent and nematode response to the chemicals.

❖ ToxPi (U.S. EPA)

The Toxicology Priority Index (ToxPi) is a 'weight of evidence' method developed by the U.S. EPA to aid in the rational prioritization of chemicals for further evaluation (Reif et al. 2010). The ToxPi score is visualized as component slices of a unit circle, where each slice represents a different component of information, such as chemical descriptors (e.g., a derived octanol/water partition coefficient or predicted % human absorption). The distance from the origin of the circle is proportional to the normalized value of the component data points in that slice (e.g., normalized assay potency) and the width of each slice, measured in radians, indicates the relative weight of the slice in the ToxPi score. When ToxPi was applied to the 309 unique ToxCast™ chemicals screened for *in vitro* activity in estrogen, androgen, and thyroid pathways, the method provided more robust conclusions by including information from different data sources than when relying on a single data source alone. In the endocrine disruptor study (Reif et al, 2010), this tool incorporated information from *in vitro* screening assays, chemical descriptors, and biological pathways to generate a prioritization rank for each chemical tested. The ToxPi procedure is flexible and may incorporate other domains of information into its score.

III.4.3.2 Prediction modeling

Predictive computational models can be used to complement experimental approaches for chemical prioritization for toxicity testing. Such approaches may not only reduce the time and cost associated with testing but could reduce or replace the need for animal testing. Historically, quantitative structure-activity relationships (QSAR) models have been used for toxicity prediction (Mohan et al. 2007). QSAR models are often limited by characteristics of the compounds used to build (or train) the model and work best when examining structurally related compounds. A broad range of compound features need to be examined to develop models with good predictive behavior. Prediction modeling algorithms are implemented in Tox21 databases (e.g., ACToR) and external commercial databases (e.g., Leadscope®, GeneGo Metacore™), and algorithms are also freely available in the R programming language and software environment (e.g., Guha 2007, Cao et al. 2008). As described below, most of these methods are currently being evaluated within Tox21.

❖ The NCGC BioPlanet of pathways

There is no one comprehensive and uniform resource that covers all known annotations of pathways or any single platform that allows integrated browsing, retrieval, and analysis of information from the many existing individual pathway resources. In response to this need, the NCGC built an integrated pathway resource that hosts information from manually curated and publicly available resources. The NCGC BioPlanet (http://www.ncgc.nih.gov/pub/bioplanet/) complements this pathway warehouse by allowing easy browsing, visualization, and analysis of the universe of pathways. The main view of the BioPlanet shows the mapping of all known human pathways on a 3D globe, where each spot represents a pathway (see **Figure III.4-6**). Selecting a pathway on the globe will place all components of the selected pathway in the detailed view window. Detailed descriptions of all genes in the selected pathway are shown below the 3D graphics. When multiple pathways are selected at the same time, the view will show all unique gene components within selected pathways. Since the coordinates of genes are

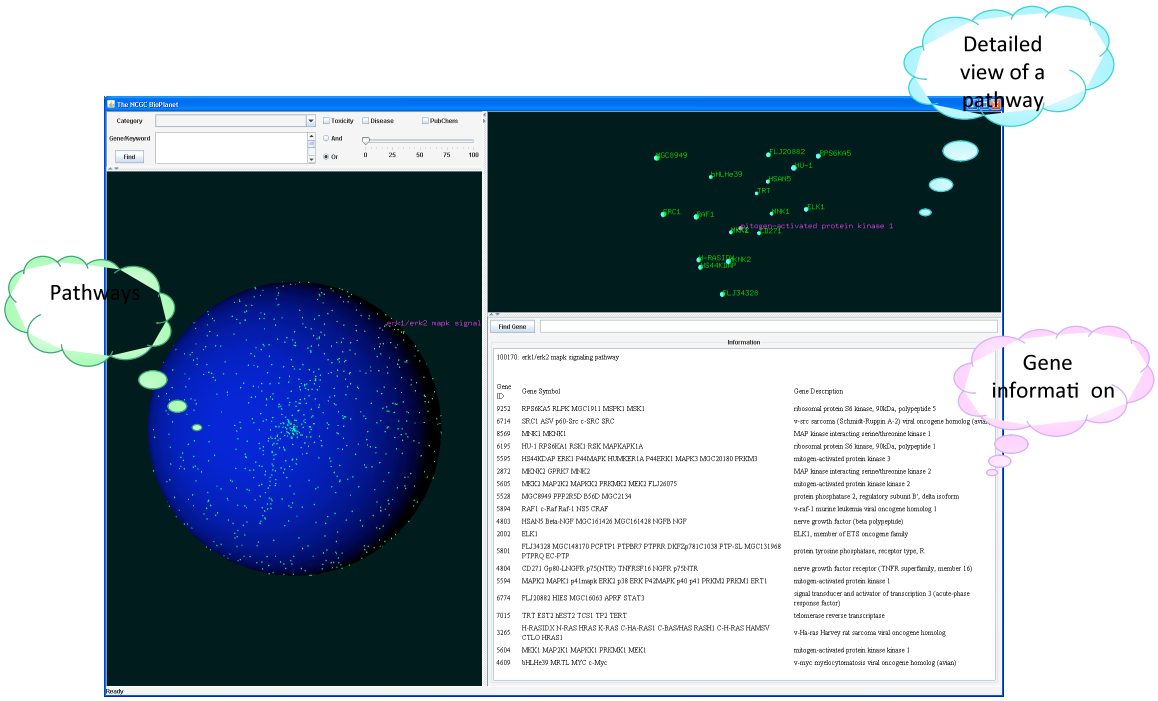


Figure III.4-6 The NCGC BioPlanet of pathways

generated globally and are not tied to any specific pathway, this allows the visualization of multiple predefined pathways as one extended pathway that better shows the interaction between different biological processes. The BioPlanet is searchable by any gene or pathway identifier, and also by disease relevance (prevalence of disease genes), toxicity relevance (occurrence of genes in toxicology literature), and availability of probing assays in PubChem.

The collection of qHTS data sets within Tox21 could be used to create a systematic view of the biological network corresponding to chemical perturbations. To achieve this goal, a set of assays needs to be selected or designed to measure targets that encompass all pathways relevant to toxicity. However, what constitutes a "toxicity pathway" is not clearly defined since our current understanding of the biological system is not sufficient for us to pinpoint the specific subset of pathways that could result in adverse health effects if disrupted. It is hoped that in the future the NCGC BioPlanet will provide a starting point for the systematic design of experiments to better understand how the biological system functions.

❖ Weighted feature significance (NCGC)

Weighted feature significance (WFS) was developed by the NCGC to predict the toxicological activity of compounds based on overrepresentation of structural features in toxic compounds (Huang et al. 2009). The WFS approach is a two-step scoring algorithm. In the first step, a Fisher's exact test is used to determine whether a given structural feature is enriched in active compounds compared to inactive compounds. Each structural feature is examined in this way, one at a time, and a p-value is calculated for every structural feature. In the second step, a WFS score is calculated according to the following equation:

$$WFS = \frac{\sum \log(p_i)}{\min(\log(p_i)) x \left(\alpha N_{C \cup M} + N_{M \cap C}\right)} \quad \text{(Eqn. 4)}$$

where p_i is the Fisher's p-value for feature i, α is the weighting factor (normally set to 1), *C* is the set of all features in the compound that were examined, *N* is the total number of features, and *M* is the set of all features present in at least one compound. As evident in Eqn. 4 the WFS score is molecule fragment-based and does not rely on whole molecule similarity for toxicity prediction, which represents a notable different from most previous methods to relate structure and activity that are available in the literature. The WFS model was trained and tested on qHTS cytotoxicity from 13 cell types (Xia et al. 2008) and caspase activation data (Huang 2008) and found to have similar predictive power when applied to mutagens or slightly better predictive power when applied to hepatotoxic compounds compared to Naïve Bayesian clustering and support vector machines (SVM) (Huang et al. 2009).

III.4.4 Toxicity Databases and Database Exploration Tools

The ability to mine historical toxicity databases is important in both testing and research endeavors and across virtually every stage of toxicological analysis. One example of the usefulness of different toxicity databases is the WormTox study (see "WormTox Modeling" above), in which results from nematode assays were compared to various data present in a rodent toxicity database (see "ToxRefDB", below). Databases also provide the information used in predictive toxicology, which is becoming increasingly important as legislative bodies require the reduction or elimination of animal toxicity testing (Yang et al. 2006). A large amount of toxicity information is publicly available through Tox21 and other government supported endeavors. However, there is still a vast array of data that is not available publicly (e.g., not released yet on public servers or available only through commercial avenues). Also, there is a range of usefulness present in some databases, since the association between chemical structure and toxic activity is not always readily available or available from different systems in data formats that are difficult to integrate into a common structure. Furthermore, different databases may incorporate results from different sources, data types, or experimental systems. For this reason, the Tox21 effort is utilizing information from a variety of different public and commercial databases and developing new databases that combine information into more useful formats for toxicity studies.

III.4.4.1 Tox21 databases

- **Chemical Effects in Biological Systems (CEBS)**

The NTP Chemical Effects in Biological Systems (CEBS) database (http://cebs.niehs.nih.gov) is a public data repository that integrates toxicity data, including qHTS data, with toxicogenomics data (Waters et al. 2008). CEBS includes information pertaining to the study design, study timeline, clinical chemistry, and pathology from different sources and is well-suited to eventually house data generated from NTP studies. In addition, CEBS links this information with relevant microarray and proteomics data and can be used to examine toxicological responses to chemicals as well as study chemical test articles, environmental agents, responses to genetic changes, and effects of physical agents. Data is available from different sources and is queried based on systemic nomenclature rather than study-specific terminology. Data queries can be based on a per-study or per-compound basis. Our goal is to house all Tox21 data in CEBS as well as well as all NTP legacy data.

❖ Aggregated Computational Toxicology Resource (ACToR)

The U.S. EPA Aggregated Computational Toxicology Resource (ACToR) is a freely distributable and open source chemical database (see http://actor.epa.gov/actor/faces/ACToRHome.jsp) that provides many types of toxicity information collected from many different sources on over 500,000 environmental chemicals in a collection of linked databases (Judson et al. 2008). Chemicals are searchable by chemical name, CAS (Chemical Abstracts Service) registry numbers, chemical structure, and other identifiers, but the database can also be browsed by assay or toxicity criteria. The data in ACToR includes information on chemical structure and *in vitro* and *in vivo* assays from over 500 publically available data collections. There are plans to include genomic and biological pathway information in the database in the future. Chemicals are linked to important toxicological parameters, namely, hazard, chronic effects, carcinogenicity, genotoxicity, developmental data, reproductive information, food safety, and exposure. Database sources include data from the U.S. federal and state governmental agencies and corresponding agencies in Canada, Europe, and Japan, as well as data produced in universities and non-governmental organizations. ACToR provides chemical activity data from toxicologically relevant sources with an organizational design similar to the design found in PubChem (see below).

One important component of ACToR is called ToxMiner™, which consists of a computer system that is able to compile, assess, and analyze high-dimensional *in vitro* and *in vivo* data for predictive models. The ToxMiner™ database is located within ACToR as a statistical tool for finding univariate associations between *in vitro* and *in vivo* data. It can also be used to produce machine learning predictive signatures.

❖ Toxicity Reference Database (ToxRefDB)

The U.S. EPA Toxicity Reference Database (ToxRefDB) is a publicly available data collection composed of data generated on hundreds of chemicals studied in almost 2,000 *in vivo* pesticide registration toxicity studies (http://actor.epa.gov/toxrefdb/faces/Home.jsp). The data in ToxRefDB are processed into structured and searchable categories that include information on study design, dosing, and treatment effects (Martin et al. 2009). This database has *in vivo* toxicity data related to rat chronic, mouse and rat cancer, rat and rabbit developmental, and rat reproductive studies and is linked to ACToR (see above). The availability of this extensive laboratory animal toxicity data provides a mechanism for validation of qHTS assay results in predictive toxicology. Data can be searched by endpoint criteria (e.g., species, in-live observations, different effects, gender) or chemical identifier (CAS registry number or chemical name). The data can be downloaded in spreadsheet format and processed in an alternative computational framework for correlating *in vitro* qHTS data and *in vivo* response.

III.4.4.2 External databases

❖ GeneGo Metacore™

Metacore™ is a manually curated database designed for analyzing experimental data in the context of pathways and networks for any high-throughput data format, including gene expression, SNPs (single nucleotide polymorphisms) and CGH (comparative genomic hybridization) arrays, proteomics, and metabolomics. Networks can be developed and visualized for human protein-compound, protein-protein, protein-RNA and protein-DNA interactions and numerous other interactions based on signaling and metabolic pathways for human, rat and mouse. Affected pathways related to compound sets can be identified and ranked for any of the pre-built pathways built from relationships described in the scientific literature. Results may be saved and exported for presentation or further analysis. The database can be queried by compound, reaction, pathway, diseases, drug, protein, or gene identifiers. The Toxicity workflow option is of particular interest for Tox21 efforts. When analyzing data in the

Toxicity workflow, pre-defined toxicity-related processes in liver, kidney, and heart are evaluated for enrichment.

❖ Leadscope® Software

Leadscope® is an interactive computer program for visualizing, exploring, and interpreting HTS data (Roberts et al. 2000). The application was originally designed as a tool to assist pharmaceutical chemists in finding drug candidates in large sets of compounds associated with biological activity data. However, the underlying features of the software also permit the exploration of qHTS data for toxicological purposes. An important component of Leadscope is the structural feature hierarchy that permits a systematic substructural analysis of compounds based on such features as amino acids, bases, benzenes, functional groups, heterocyclic, and many other categories. Sets of active compounds with particular structural features that are associated with toxicological data can be found by exploring the extensive database of chemical features and properties in the Leadscope Enterprise™ database. Features include generalized atom pairs, molecular fingerprints, chemical substructure, two- and three-dimensional descriptors, physiochemical properties, partial atomic charges, and other descriptors (Roberts et al. 2000). Chemical structural features that are statistically correlated with toxicological activities can be found based on QSAR, similarity searching, and clustering. Statistical correlations are presented as histogram bars and scatter plot cells, which are based on the difference between the mean activity of a compound set with a particular feature and the entire compound set. Datasets containing information for both active and inactive compounds can help to find individual features or combinations of features that are highly correlated with activity by exploration of the frequency of the structural features present in a compound set as well as statistical measures of correlation activity.

In addition, a new multi-tier client/server/database system is being developed through a phase I NIH SBIR contract, the purpose of which is to support the integration of toxicological relevant data from diverse sources. These data will be augmented with online toxicity predictive model results from several sources including Leadscope's and MultiCASE's QSAR models and Derek structural alerts from Lhasa Limited. In addition, a number of simple tools are proposed to help scientists make sense of the information. Methods to help prioritize chemicals for more extensive testing based on similar profiles, such as from qHTS or gene expression data, to a reference chemical are being explored. Also, tools to support understanding mechanisms of action, human relevant predictions, and NICEATM validation studies are also being investigated. An important aspect of the proposed software is that it should be available through web browsers on different operating systems and it should support scientists with diverse experience and training as well as integrating with current NIEHS' workflows and systems.

❖ Comparative Toxicogenomics Database (MDIBL)

The Comparative Toxicogenomics Database (CTD) is a freely available database (see http://ctd.mdibl.org/) used to help elucidate the mechanisms by which environmental chemicals affect human health. The database is manually curated, containing literature-based relationships from over 77,000 chemical-gene interactions, 2500 chemical-disease interactions, and 350,000 gene-disease associations (Davis et al. 2009). The integration of these three basic sets of relationships provides insight into chemical-gene-disease networks affecting human disease. The information within the CTD can be queried within a keyword search box by a single chemical, disease, gene, Gene Ontology (GO) terms, organism, or references. The CTD can also be searched in batch form. Database query results can be viewed from the perspective of a chemical, gene or disease, each of which is hyperlinked to more detailed pages. Networks resulting from all searches can be downloaded into CSV (comma-separated

values), TSV (tab-separated values) or XML (extensible markup language) format for use in other analysis platforms.

❖ **PubChem (NCBI)**

PubChem, a component of the NIH Molecular Libraries Roadmap Initiative found at http://pubchem.ncbi.nlm.nih.gov/ , is a chemical database for molecules with fewer than 1000 atoms and 1000 chemical bonds (Austin et al. 2004). PubChem is comprised of three primary databases, namely Substances, Compound, and BioAssay. The BioAssay database contains results from bioactivity assays of over 1500 HTS programs and represents the largest database resource for integrating chemical structure and data generated using HTS assays. Three main data types are catalogued in PubChem: substances are indexed by substance identifier (SID), compounds are indexed by compound identifier (CID), and bioassays are indexed by assay identifier (AID). In this organization, a substance is defined as a single chemical entity submitted by one data source, a compound refers to chemical structures, and a bioassay represents specific test data associated with one or more substances (Judson et al. 2008). Tox21 data generated at the NCGC have been and will continue to be made available to the scientific community via this database.

III.4.5 Tox21-related Publications

Huang R, Southall N, Xia M, Cho M-H, Jadhav A, Nguyen, et al. 2009. Weighted Feature Significance (WFS): a simple, interpretable model of compound toxicity based on the statistical enrichment of structural features. Toxicol Sci 112(2):385-93.

Huang R, Southall N, Xia M, Cho M-H, Jadhav A, Nguyen, et al. 2009. Weighted Feature Significance (WFS): a simple, interpretable model of compound toxicity based on the statistical enrichment of structural features. Toxicol Sci 112(2):385-93.

Huang R, Xia M, Cho M, Sakamuru S, Shinn P, K Houck, et al. 2010. Chemical genomics profiling of environmental chemicals on human nuclear receptors. Environmental Health Perspectives, submitted.

Parham F, Austin C, Southall N, Xia M, Tice R, Portier C. 2009. Dose-response modeling of high-throughput screening data. J Biomol Screen 14(10):1216-27.

Reif D, Martin M, Tan S, Houck K, Judson R, Richard A, et al. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast™ data. Environ Health Perspect, published 08 Sep 2010, doi:10.1289/ehp.1002180.

III.4.6 References

Austin CP, Brady LS, Insel TR, Collins FS. 2004. NIH molecular libraries initiative. Science 306(5699):1138-1139.

Cao Y, Charisi A, Cheng L-C, Jiang T, Girke T. 2008. ChemmineR: a compound mining framework for R. Bioinformatics 24(15):1733-1734.

Collins FS, Gray GM, Bucher JR. 2008. Toxicology. Transforming environmental health protection. Science 319(5865):906-907.

Davis AP, Murphy CG, Saraceni-Richards CA, Rosenstein MC, Wiegers TC, Mattingly CJ. 2009. Comparative Toxicogenomics database: a knowledgebase and discovery tool for chemical-gene-disease networks. Nucleic Acid Res 37:D786-D792.

Guha R. 2007. Chemical informatics functionality in R. J Stat Software 18(5):1-16.

Hill AV. 1910. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J Physiol 40:4-7.

Huang R, Southall N, Cho M-H, Xia M, Inglese J, Austin CP. 2008. Characterization of diversity in toxicity mechanism using *in vitro* cytotoxicity assays in quantitative high throughput screening. Chem Res Toxicol 21(3):659-667.

Huang R, Southall N, Xia M, Cho M-H, Jadhav A, Nguyen D-T, et al. 2009. Weighted feature significance: a simple, interpretable model of compound toxicity based on the statistical enrichment of structural features. Tox Sci 112(2):385-393.

Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, et al. 2006. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U.S.A. 103(31):11473-11478.

Inglese J, Johnson RL, Simeonov A, Xia M, Zheng W, Austin CP, et al. 2007. High-throughput screening assays for the identification of chemical probes. Nat Chem Biol 3(8):466-479.

Judson R, Richard A, Dix D, Houck K, Elloumi F, Martin M, et al. 2008. ACToR—Aggregated computational toxicology resource. Toxicol Appl Pharmacol 233(1):7-13.

Judson RS, Houch KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118(4):485-492.

Martin MT, Judson RS, Reif DM, Kavlock RJ, Dix DJ. 2009. Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef database. Environ Health Perspect 117(3):1302-1314.

Mohan CG, Ghandi T, Garg D, Shinde R. 2007. Computer-assisted methods in chemical toxicity prediction. Mini Rev Med Chem 7(5):499-507.

Muir DC, Howard PH. 2006. Are there other persistent organic pollutants? A challenge for environmental chemists. Environ Sci Technol 40(23):7157-7166.

Parham F, Austin C, Southall N, Huang R, Tice R, Portier C. 2009. Concentration-response modeling of high-throughput screening data. J Biomol Screen 14(10):1216-1227.

Reif D, Martin M, Tan S, Houck K, Judson R, Richard A, et al. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ Health Perspect, published 08 Sep 2010, doi:10.1289/ehp.1002180.

Ritz C, Streibig JC. 2005. Bioassay analysis using R. J Stat Software 12(5):1-22.

Roberts G, Myatt GL, Johnson WP, Cross KP, Blower PE. 2000. Leadscope: Software for exploring large sets of screening data. J Chem Inf Comput Sci 40(6):1302-1314.

Shockley K, Kissling G, Huang R, Xia M, Austin C, Tice R. 2010. A statistical framework for analyzing quantitative high-throughput screening data (qHTS). The Toxicologist 114(1): 42.

Shukla SJ, Huang R, Austin CP, Xia M. 2010. Foundation review: The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform. Drug Discov Today doi:10.1016/j.drudis.2010.07.007.

Waters M, Stasiewicz S, Merrick BA, Tomer K, Bushel P, Paules R, et al. 2008. CEBS – Chemical effects in biological systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucl Acid Res 36:D892-D900.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho M-H, et al. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect 116(3):284-291.

Yang C, Benz RD, Cheeseman MA. 2006. Landscape of current toxicity databases and database standards. Curr Opin Drug Discov Devel 9(1):124-133.

Zhang J-H, Chung TWY, Oldenburg KR. 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screening 4(2):67-73.

III.5 Tox21 Targeted Testing Working Group

This page left intentionally blank

III.5 Tox21 Targeted Testing Working Group

Michael DeVito, Ph.D.
NTP Toxicology Branch

Tox21 Targeted Testing Working Group
R. Daniel Benz, Ph.D. (U.S. FDA CDER)*
Kevin Crofton, Ph.D. (U.S. EPA NHEERL)*
Michael DeVito, Ph.D. (NTP TB)
James Inglese, Ph.D. (NHGRI NCGC)*
Scott Masten, Ph.D. (NTP Office of Chemical Nomination and Selection)*
B. Alex Merrick, Ph.D. (BSB)
** = Co-Chairs*

NTP Scientists
Scott Auerbach, Ph.D. DABT (BSB)
Kristine Witt, M.S. (BSB)

Collaborating Scientists
Christopher Corton, Ph.D. (EPA NHEERL)
Richard Judson, Ph.D. (U.S. EPA NCCT)
Imran Shah, Ph.D. (U.S. EPA NCCT)
Keith Houck, Ph.D. (U.S. EPA NCCT)

III.5.1 Background and Rationale

Long-term animal tests remain the gold standard for predicting the toxicity of environmental chemicals. One problem faced by regulatory agencies and the NTP is that there are thousands of chemicals that require testing. The sheer number of chemicals needing testing is beyond our capacity and necessitates the development of prioritization methods. Presently at NTP, prioritization is done in an ad hoc manner through an external and internal nomination process. An alternative approach to toxicity testing is one that uses the results of *in vitro* HTS as a means of prioritization. Ideally, HTS results would not only aid in the prioritization of which chemicals to test, but would also streamline testing by prioritizing which tests would be of most value for higher priority chemicals. It might be possible to envision this "targeted testing" approach as a means of significantly reducing animal usage by only using animals for those chemicals and endpoints that are most critical.

The concept of toxicity pathways is critical in the implementation of toxicity testing in the 21st century. While defined in the NAS Report (2007), the identification of toxicity pathways presents considerable challenges and even the number of these pathways remains elusive. Once identified, there are a variety of *in vitro* approaches, using different technologies and cell types, available to assess chemical alterations of these pathways. The screening of hundreds of assays with hundreds to thousands of chemicals results in large data sets requiring significant bioinformatic efforts to develop both predictive models and prioritization schemes.

Among the four working groups within the Tox21 community, the targeted testing working group designs studies to evaluate the relevance of prediction models and prioritization schemes developed from Tox21 data. The U.S. EPA NCCT has developed a statistical model that uses ToxCast™ Phase I data to predict nongenotoxic rodent liver tumorigens (Judson et al. 2010). Of the 309 individual chemicals tested in ToxCast™ Phase I, 248 have 2-year rat chronic/cancer bioassay data available in ToxRefDB. Twenty-one of these chemicals induced hepatocyte-derived liver tumors. The relationship between the results from ToxCast™ Phase I and the likelihood of developing rat liver tumors was evaluated by testing univariate associations of all *in vitro* assays and gene perturbation scores against all *in vivo* rodent liver end points. For the 21 chemicals, this approach identified a total of seven *in vitro* assays where a positive response had a significant association with the presence of rat liver tumors (Fisher's exact test p-value < 0.01) (**Table III.5-1**). We fully appreciate that the ultimate goal of Tox21 is the prediction of human health effects. However, given our current lack of human relevant toxicological data, our short-term goal is to evaluate the relevance of *in vitro* prediction models to whichever *in vivo* model is most appropriate, regardless of the species involved.

Table III.5-1 Significant univariate associations between ToxCast™ assays and rat liver tumors

Assay name	Species	Gene	Cell	Assay type	Pathway
ATG_PPARg_TRANS	Human	PPARγ	HepG2	Reporter gene	PPARγ
ATG_PPARa_TRANS	Human	PPARα	HepG2	Reporter gene	PPARα
NCGC_AR_Antagonist	Human	Androgen Receptor	HEK293H	Reporter gene	Androgen Receptor
NCGC_PPARa_Agonist	Human	PPARα	HEK293H	Reporter gene	PPARα
CLZD_HMGCS2_48	Human	HMGCS2	Primary hepatocytes	qNPA	PPARα
BSK_SM3C_MCP1_up	Human	CCL2	HUVEC, Primary Human Vascular Smooth Muscle Cells	ELISA	chemokine (C-C motif) ligand 2
CLM_OxidativeStress_24hr	Human	H2AFX-P	HepG2	Fluorescence	Oxidative Stress

Abbreviations: AR = androgen receptor; CCL2 - Chemokine (C-C motif) ligand 2; PPAR – peroxisome proliferating antigen receptor; qNPA = quantitative nuclease protection assay

When this model is applied to the full ToxCast™ Phase I data set, 69 chemicals are predicted to be rat liver carcinogens.

The U.S. EPA NCCT model predicts that non-genotoxic chemicals that activate the PPARγ signaling pathway and one or more of the following pathways: PPARα activation, cytokine CCL2 up-regulation (CCL2), androgen antagonism (AA), or oxidative stress (OS), have a significantly increased likelihood for inducing rat liver tumors when compared to non-genotoxic chemicals activating none or only one of these processes. This model provides interesting and novel results. The importance of PPAR in rodent liver carcinogenesis is well documented (Klaunig et al. 2003); however, most of the data describes the

importance of PPARα and much less attention has been paid to the role of PPARγ in rodent hepatocarcinogenicity. CCL2 has been implicated in up-regulation of bile acids, liver injury, and fibrogenesis (Ramm et al. 2009). The AR and levels of androgens have been associated with hepatocellular carcinomas (HCCs) (Kalra et al. 2008). The Cellumin Oxidative stress assay uses a fluorescent probe to measure changes in H2AFX phosphorylation.

The phosphorylated form of this enzyme is a double stranded DNA break repair enzyme. While oxidative stress and DNA damage are clearly involved in carcinogenicity, these chemicals are negative when tested in standard genotoxicity assays (ToxRefDB 2010).

There are several challenges in the interpretation of this predictive model. One of the most important is that of the *in vitro* assays use human tissue derived cells while the *in vivo* toxicity data is based on experimental animals. Thus, the prediction model uses human *in vitro* data to predict rodent *in vivo* data and the validity of this approach has not been evaluated. It is likely that this approach will have limitations. For example, in rodents, it is clear that activators of the constitutive androstane receptor (CAR) induces liver tumors, yet CAR activation is not one of the pathways indentified as predicting rat liver tumors. One possible explanation for this is that there are significant differences in the CAR ligand binding domain between humans and rats (Timsit and Negishi 2007). The impact of these species differences on our approach to developing predictive models needs to be evaluated. Another issue is that while it is possible to hypothesize the role for each of these pathways in the development of rat liver tumors, there is limited *in vivo* evidence that these pathways are activated at carcinogenic doses of these chemicals. Finally, metabolism is not included in the *in vitro* assays and this lack might lead to spurious correlations. As part of its goal to evaluate the relevance of prediction models and prioritization schemes, the targeted testing work group designed limited *in vivo* studies to determine the concordance between the *in vitro* findings to *in vivo* responses.

The present studies will evaluate the *in vivo* dose-response relationships for signature events predicated by the U.S. EPA NCCT model for chemicals that cause or are predicted to cause liver cancer in the rat. The proposed short-term *in vivo* studies provide a method for qualitatively evaluating the concordance between the *in vitro* screening assays and *in vivo* responses. Anchoring the pathway signatures to pathological findings *in vivo* is a critical goal of this study.

III.5.2 Key Issue, Hypothesis Tested, or Problem Addressed

Key Issue: Are the pathways that are perturbed by a chemical *in vitro* also perturbed *in vivo*, and in the tissues of interest?

Hypothesis 1: *In vitro* activation of PPARγ/CCL2/AA/OS/PPARα is highly predictive of the corresponding activation *in vivo*, at some dose level.

Hypothesis 2: Only at doses for which at least two of these pathways or processes are activated will liver tumors be observed in a 2-year rat study.

III.5.3 Approach

In vivo studies in rats will be used to assess the relevance of this prediction model. Rats will be treated short term with ToxCast™ Phase I chemicals predicted by the model to induce rat liver tumors. To evaluate whether the *in vitro* assay results occur *in vivo*, we need to examine hepatic markers consistent with those used in the *in vitro* studies**. Table III.5-2** presents the *in vivo* endpoint with the corresponding *in vitro* assay endpoint. The initial exposure paradigm is four daily doses with the study terminated 4

hours after the last dose. This exposure paradigm was chosen based on consideration that most *in vitro* assays were 24-48 hours of incubation and that alterations in gene expression for hepatic nuclear receptors peak around 3-4 days of *in vivo* treatment Sampling at 4 hours after the last treatment is the typical protocol at NTP for the Comet Assay, an assay for detecting DNA damage in single cells.

Table III.5-2 *In vivo* endpoints and corresponding *in vitro* assay

In vivo Endpoints	*In vitro* Assay[a]	Justification
Affymetrix GeneChip Rat RAE230 2.0	PPARγ / CCL2 / AA / OS/PPARα	The *in vitro* assays are thought to be markers for pathway activation and the arrays are the most efficient method to assay all of these pathways. Pathway analysis will be done on PPAR, oxidative stress, and androgen receptor.
Immunohistochemistry for phosphorylated-gamma-H2AX	Cellumin Oxidative stress assay	The *in vitro* assay is an imaging assay that measures the amount of phosphorylated gamma H2AX. Thus, a corresponding assay would be to measure phosphorylated H2AX protein.
Genotoxicity Comet Assay on liver tissue (traditional and oxidative damage specific comet assay)	Cellumin Oxidative stress assay	Since the Cellumin oxidative stress assay is really a measure of DNA repair, the design team thought it would be of value to reassess the genetic toxicity of these chemicals with newer methods.
RT-PCR for HMG-CoA synthase 2 (HMGCS2)	CellzDirect hepatocyte assay PPARα	The PPARα assay was the induction of HMGCS2 in human hepatocytes after 48 hours of exposure to the test chemical. This is a direct correlate for that assay *in vivo*.
RT-PCR for hepatic medium chain acyl CoA dehydrogenase and phosphoenolpyruvate carboxykinase mRNA	PPARγ - transactivation assay	These two genes are directly regulated by PPAR and induction of these genes is a close correspondence to the *in vitro* assay for PPARγ activation.
ELISA for CCL2	CCL2 protein down regulation	This is a direct correspondence to the *in vitro* assay.
Clinical chemistry on blood for glucose, cholesterol, and triglycerides; high density lipoprotein; low density lipoprotein; alanine aminotransferase; sorbitol dehydrogenase	PPARγ - transactivation assay	PPARγ agonists alter glucose and lipid concentrations in rodent serum in short-term assays.

Abbreviations: AA = androgen receptor antagonist; CCL2 = Chemokine (C-C motif) ligand 2; ELISA = enzyme-linked immunosorbent assay; HMG-CoA = 3-hydroxy-3-methylglutaryl-Coenzyme A; OS = oxidative stress; PPAR = peroxisome proliferating antigen receptor; RT-PCR = real-time polymerase chain reaction.

III.5.4 Summary of Previous Study Data

The chemicals to be evaluated in this study are part of the ToxCast™ Phase I project. As such, each chemical has HTS assay results and the effects of these chemicals in the relevant HTS assays are listed in **Table III.5-3**. In addition, these chemicals have two-year bioassay data in rats.

Table III.5-3 Chemicals and *in vivo* and *in vitro* results*

CHEMICAL	Tumor Results		ToxCast™ Results				
	Rat Liver Tumors** (%)	High Dose (mg/kg)***	PPARγ	OS	AA	HMGCS2	CCL2
Acetoclor	+ (7)	250	+	+	+	+	-
Dimethenamid	+(12)	109	+	+	+	-	-
Lactofen	+(9)	79	+	-	+	+	-
PFOA	+ (13)	300	+	-	-	+	-
PFOS	+ (9)	100	-	-	-	-	-
2,5-Pyridinedi-carboxylic acid di-n-propyl ester	+(30)	1000	+	-	-	-	+
Simazine	+	63	+	-	-	-	+
Carbaryl	+ (6)	485	+	+	-	-	-
Bisphenol A	-	1000	+	-	+	-	-
Fludioxonil	-	121	+	+	-	-	-
Flusilazole	-	13	+	-	+	-	-
Triclosan	-	1000	-	-	-	-	-

Abbreviations: AA - Androgen Receptor antagonist; CCL2 = Chemokine (C-C motif) ligand 2; HMGCS2 = 3-hydroxy-3-methylglutaryl-CoA synthase 2 (mitochondrial); OS = oxidative stress; PPARγ = peroxisome proliferator-activated receptor-γ

*- Data for the tumor results are available on ToxRefDB (2010) and the ToxCast™ results are available at http://epa.gov/ncct/toxcast/data_sets.html

** a "+" indicates positive in the two-year bioassay for liver tumors or ToxCast™ assay and a "-" indicates no treatment related liver tumors were found in the two-year bioassay or treatment related activity in the ToxCast™ assays.

***Highest dose administered in two-year bioassay that was positive for rat liver tumors or if the chemical was did not induce hepatic tumors, the highest dose administered in the bioassay.

III.5.5 Critical Decision Points

We plan to test a number of chemicals in this study in a phased approach. In Phase I, we will begin by evaluating 12 chemicals (**Table III.5-3**) in male Harlan Sprague-Dawley rats. These chemicals were chosen from the 69 ToxCast™ chemicals the model predicts are rat liver carcinogens. In addition, these chemicals were chosen based on coordination with the Virtual Liver project at the U.S. EPA NCCT (see http://www.epa.gov/ncct/virtual_liver/). If we observe a concordance between the *in vitro* signatures and the *in vivo* results, we will continue with another set of chemicals and again assess the results prior to any additional studies. If there is a lack of concordance between the *in vitro* and *in vivo* results, the study design team will assess the results and present a summary of their analyses and recommendations on further studies to the Tox21 community. If there is concordance between the *in vitro* and *in vivo* responses, a subset of chemicals will be evaluated for dose response and time course. These studies will include tissue dosimetry in an effort to relate tissue concentrations to *in vitro* concentrations.

III.5.6 Significance

The success of *in vitro* screening and toxicity testing depends on the development of prediction models. Initial attempts to build these models are based on the statistical associations between HTS results and rodent toxicity databases. The focus of the Targeted Testing working group is to evaluate the relevance of these prediction models. The ongoing studies described will assess the *in vivo* relevance of the U.S. EPA NCCT prediction model. These efforts should also provide insight into the qualitative and quantitative relationship between the *in vitro* and *in vivo* assays.

III.5.7 Future Directions

The present study only evaluates a model that predicts rat liver tumors. As other models are developed, there might be a need to evaluate them as well. Initial discussions have begun on developing a targeted testing project for endocrine disrupting chemicals and reproductive and developmental toxicants.

III.5.8 References

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118(4):485-492.

Kalra M, Mayes J, Assefa S, Kaul AK, Kaul R. 2008. Role of sex steroid receptors in pathobiology of hepatocellular carcinoma. World J Gastroenterol 14(39):5945-61.

Klaunig JE, Babich MA, Baetcke KP, Cook JC, Corton JC, David RM, et al. 2003. PPARalpha agonist-induced rodent tumors: modes of action and human relevance. Crit Rev Toxicol 33(6):655-780.

Ramm GA, Shepherd RW, Hoskins AC, Greco SA, Ney AD, Pereira TN, et al. 2009. Fibrogenesis in pediatric cholestatic liver disease: role of taurocholate and hepatocyte-derived monocyte chemotaxis protein-1 in hepatic stellate cell recruitment. Hepatology 49(2):533-44.

Timsit YE, Negishi M. 2007. CAR and PXR: the xenobiotic-sensing receptors. Steroids 72(3):231-46.

ToxCast™ Data Sets & Published Research (August 3, 2010) http://epa.gov/ncct/toxcast/data_sets.html

ToxRefDB: Toxicity Reference Database: (August 3, 2010) http://actor.epa.gov/toxrefdb/faces/Home.jsp

IV. Tox21 Activities

IV.1 Introduction to Tox21 Activities

This page left intentionally blank

IV.1 Introduction to Tox21 Activities

Raymond T. Tice, Ph.D.
Chief, Biomolecular Screening Branch

As stated earlier, the goals of Tox 21 are to:

- research, develop, validate, and translate innovative chemical testing methods that characterize toxicity pathways
- identify chemicals, assays, informatics analyses, and targeted testing needed for the innovative testing methods
- prioritize chemicals for more extensive toxicological evaluation
- identify mechanisms of chemically-induced biological activity in order to characterize toxicity pathways, facilitate cross-species extrapolation, and provide input to models for low-dose extrapolation
- develop predictive models for biological response in humans

while the related goals of the BSB are to:

- carry out the NTP automated screening assays with *C. elegans*
- develop research and testing activities in high and medium throughput screening assays for rapid detection of biological activities of significance to toxicology and carcinogenesis
- develop analysis tools and approaches to allow an integrated assessment of HTS endpoints and associations with findings from traditional toxicology and cancer models
- develop assays and approaches to understand the genetic and epigenetic bases for differences in susceptibility

This part of the review presents major activities initiated by the BSB/Tox21, in addition to the ones already described in the Tox21 Working Group presentations, to achieve these goals; these are activities that are not connected with the Host Susceptibility activities reviewed at the Board meeting in December 2009.

IV.1.1 NTP *Caenorhabditis elegans* Screening Facility (Presenter: Jonathan Freedman, Ph.D.; Laboratory of Toxicology and Pharmacology, Environmental Toxicology Program)

The NTP *C. elegans* Screening Facility, also known as the WormTox group, is led by Dr. Jonathan Freedman, the Principle Investigator of the Comparative Genomics Group. The *C. elegans* Screening Facility was established in 2003 (and became part of the BSB when it was established in 2007), with the goal to develop toxicological assays using the nematode *Caenorhabditis elegans. C. elegans* has been used as a biological model for human disease for decades, which has led to numerous discoveries in genetics, development, and neurobiology. Many pathways known to govern responses to chemical

insults are highly conserved from *C. elegans* to mammals. This suggests that chemicals that elicit a toxicological response in *C. elegans* are likely to affect humans. A number of phenotypes have been observed in *C. elegans* after chemical exposures including changes in behavior, morphology, and gene expression. Many of these phenotypes are correlated with lethality in traditional rodent models, further supporting the utility of *C. elegans* in predicting mammalian toxicity. WormTox evaluates the utility of these phenotypes as components of medium- and high-throughput screening assays. Four toxicological assays have been developed to characterize the effects of chemicals on *C. elegans* – feeding, growth, reproduction, and locomotion. To date, WormTox has screened over 1,700 compounds in one or more assays including two chemical libraries: the NTP's first 1,408 and the U. S. EPA's ToxCast™ 320. In addition to screening, WormTox has performed assays to address a number of basic research questions in collaboration with extramural and intramural investigators. Future plans for WormTox include the development of new assays and comparative toxicological analyses. Currently WormTox is generating a collection of GFP-expressing, chemical- and stress-responsive transgenic strains of *C. elegans.* These strains will be used to rapidly assess the effects of toxicants on the activity of targeted pathways and the transcription of individual genes. A high priority goal of WormTox is to compare *C. elegans* ToxCast™ toxicity data to that of other organisms, including rodents and zebrafish, and *in vitro* data generated by other ToxCast™ members. This analysis is an important step in evaluating the ability of toxicological responses in *C. elegans* to predict toxicity in higher organisms.

IV.1.2 Probing Mechanisms of Inter-individual Susceptibility to Toxicants with Population-based Experimental Approaches (Presenter: Ivan Rusyn, M.D., Ph.D.; Director Laboratory of Environmental Genomics, Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill)

Early in the NTP collaboration with the NCGC, Drs. Chris Austin (NCGC), Jef French (NTP) and I discussed the possibility that the NCGC qHTS format could be used to evaluate differences in inter-individual susceptibility to toxic compounds and that such data might be useful for understanding the range of individual variability in the rodent or human population, as well as for potentially being able to provide valuable insights into the molecular mechanisms for genetically-determined variability in responses. We envisioned a qHTS project where, rather than screening thousand of compounds against the same target cell population, cells from all mouse strains or from hundreds to thousands of individuals would be screened for differential sensitivity to specific compounds selected on the basis of known toxicity. We recognized the limitations of this approach (e.g., *in vivo* variation in sensitivity might not equal *in vitro* variation in sensitivity, and vice versa; primary cells would be preferable but would likely not be available and/or amenable to qHTS) but concluded that a pilot study using densely genotyped rodent or human cells in order to link genes to any differential sensitivity detected would be valuable.

In response to this discussion, Dr. French established a project to produce cell lines from densely genotyped mouse strains in order to evaluate the mechanistic basis for inter-strain *in vivo* differences in susceptibility to specific compounds. This project is based on the hypothesis that the results obtained from the qHTS screening of these cell lines for differential sensitivity to a broad spectrum of toxic compounds could be used to identify genes associated with differences in sensitivity, and that this knowledge could be used to design *in vivo* mouse strain studies that would evaluate the relevance of the *in vitro* data to the expression of disease *in vivo*. The basis for this approach, described here but not presented, is that classic inbred mouse strains have been used as models in studies of human disease and health effects due to environmental exposures for decades. The genetic basis for toxic responses in both species is highly conserved based upon homology of phase 1, 2, and 3 metabolic systems for detoxification and DNA repair systems, although conservation of allelic variation and gene expression

pathways is largely unknown. Use of array platforms to evaluate single nucleotide polymorphisms (SNPs), copy number variants (CNV), and methylated cytosine variation and changes in gene (transcript) expression in response to toxicant exposures in both species will greatly enhance the value of mouse models in the field of toxicology. In addition, emphasis is being placed on the development of high throughput cell-based assay systems that can be used in biomolecular screens. Studies demonstrating heritable differences in cDNA expression as well as changes in gene expression as the result of toxicant exposure have been made using human lymphoblastic cell lines (LBCL) derived from B-lymphocytes. No equivalent mouse LBCL resource exists at the present time. Moreover, lack of genetic variation in the laboratory-derived inbred mouse strains makes comparison of allelic variants in areas of the mouse genome identical by descent between mouse and human difficult. We are presently creating a set of 35 mouse lymphoblastic cell lines by transfection of the human receptor complex and using the same Epstein Barr virus transformation techniques used to derive the human LBCLs. Selection of the mouse strains will include the eight parental inbred lines used to create the highly diverse Collaborative Cross recombinant inbred mouse resource, thus increasing the number of genomic variants available for observation of allelic variation and increasing the chance for observing mouse-human comparisons. These mouse LBCLs and available mouse embryonic fibroblasts (MEFs) will be used to integrate and compare host-susceptibility responses following screening in qHTS. Identification of genes and their allelic variants associated with responses to xenobiotic exposures in mouse models will aid in the identification of orthologous genes in genetically diverse human populations associated with the same or similar outcomes in response to exposures and, ultimately, will aid in the assessment of potential risk due to these exposures. The integration of the Host Susceptibility Branch into BSB is critical to this mission.

Parallel to the mouse strain studies implemented by the Host Susceptibility Branch, now a Group within the BSB, we established a research collaboration with Dr. Ivan Rusyn (University of North Carolina, Chapel Hill) in order to conduct a pilot study at the NCGC to evaluate for differential sensitivity in densely genotyped human cells. Dr. Rusyn's laboratory was conducting *in vitro* studies to evaluate differential sensitivity among a genetically defined panel of human lymphoblastoid cell lines, using a small set of compounds and evaluating the same endpoints for cytotoxicity and apoptosis that had been used at the NCGC in Tox21 Phase I. The results of this collaboration were the screening of lymphoblast cell lines from 27 Centre d'Etude du Polymorphisme Humain (CEPH) trios assembled by the HapMap Consortium with 240 compounds (a selected subset of the NTP 1408 compound library). Caspase 3/7 activity, a marker of apoptosis, and intracellular ATP production, a measure of cell viability were the endpoints evaluated. Dr. Rusyn will present the results of this study.

IV.1.3 Mining the NTP Tissue Archives for Gene Signatures (Presenter: B. Alex Merrick, Ph.D.; Biomolecular Screening Branch)

NTP archival tissues represent a significant and underutilized resource, so we have initiated pilot studies to evaluate the extent to which gene expression signatures can be reliably derived from the molecular analysis of tissue samples collected from the laboratory animals used in NTP's toxicological studies and stored as formalin fixed, paraffin embedded (FFPE) tissues in the NTP archives. Signature expression profiles are critical sets of altered transcripts or proteins that distinguish toxicity and disease from a comparable normal state. Development of such signatures would further our understanding of pathological changes, mechanisms, and critical pathways in agent-induced toxicity. Further, such signatures could contribute to the identification of useful targets for *in vitro* assays, to an evaluation of the correlation between *in vitro* test results and *in vivo* toxicological outcomes, and to the development of predictive models of toxicity. NTP toxicology studies sometimes involve storing frozen tissues but not always so that extraction of full-length RNA for later profiling may be limited. Assuming less than

optimal RNA extracted from FFPE tissue can be overcome, molecular analyses on archival paraffin block tissues would greatly expand our ability to link gene expression changes with disease outcomes while leveraging the considerable expense already invested in NTP toxicological studies. The technologies being evaluated include quantitative polymerase chain reaction (qPCR), various forms of NextGen Sequencing, and quantitative Nuclease Protection Assay (qNPA). We are also engaged in focused collaborative studies with Dr. Michael Waalkes (Chief, NTP Laboratories Branch) to relate epigenetic changes in methylated DNA with transcript expression changes by NextGen sequencing approaches and with the NTP laboratory of Dr. Jean Harry to evaluate an antibody array proteomics platform for its use with NTP archived tissues.

Related to the goal of developing analysis tools and approaches to allow an integrated assessment of HTS endpoints and associations with findings from traditional toxicology and cancer models, in September, 2010, the NTP acquired DrugMatrix®, a toxicogenomics reference database, the accompanying extensive frozen tissue archives, and the informatics system. NTP acquired this resource to expand our ability to develop predictive models for toxicological effects based on gene signatures, to provide additional tools for linking *in vitro* data to *in vivo* gene signatures and disease outcomes, and to provide additional tissue samples for NextGen-based investigations.

IV.1.4 A Bioinformatics-Based Approach to Identifying Assays That Query Human Health Effects (Scott Auerbach, Ph.D., DABT; Biomolecular Screening Branch)

The goal of this project is to create meta-database that relates genes, pathways, and biological processes to human disease and subsequently to identify the chemical genomic space within these relationships that can be exploited to query the effects of chemicals on molecular processes related to human disease. The first step is to merge data from a large number of disease/genome databases including the Comparative Toxicogenomics Database Phenopedia, Human Genome Association Database, OMIM, GeneCards, Entrez Gene, CoPub, KEGG Disease, and MedGene. These databases procure information from a number of resources including literature mining, genetic studies, and functional genomics studies. Disease:gene relationships found in these databases are cataloged and a weighted voting approach is then used to identify a rank list of genes for each disease of concern. It is then determined if the identified disease genes possess a protein domain that would provide an interaction interface for an environmental agent. Disease genes that fit this category become a priority for assay identification/generation. This approach depends in a large part on literature mining and therefore publication bias can potentially inflate the rank of a gene. In addition, understudied/underpublished disease:gene relationships may lead to a deflate rank of potentially important genes. This presentation will describe this approach using obesity and diabetes as examples.

IV.1.5 The Mouse Methylome Project (Presenter, Dr. John (Jef) French; Host Susceptibility Group, Biomolecular Screening Branch)

An individual's response to exposure-related toxicity and concomitant disease is influenced at the genome level by genetic, epigenetic, and gene-gene interactions (**intrinsic factors**) and interaction with the environment (**extrinsic factors**). Individual DNA sequence variation does not account for all of the heritability for susceptibility to toxicity and diseases such as asthma, cancer, diabetes, etc. An intrinsic factor that quantitative and molecular geneticists believe is the basis for the observed "missing heritability" is the **methylome** an individual's genome wide methylated CpG sequence pattern. The methylome (a component of the epigenome) may be the major epigenetic modifier of the susceptibility to cancer and other chemical exposure related diseases. Presently, there is no mouse reference database for the methylome akin to the NTP/Perlegen DNA sequence data of 15 commonly used inbred

strains (plus the C57BL/6 reference strain). The DNA sequence data has significantly increased our knowledge of the genomic structure of the inbred mouse and it has provided the basis for imputation of the haplotype structure of more than 90 inbred strains used in biological research. The absence of a methylome reference database for the mouse significantly handicaps our knowledge and understanding of the mouse model in toxicology and environmentally related diseases and designing and conducting hypothesis based genetic and epigenetic research studies to understand the associated mechanisms.

Two high content technologies have been recently developed that (1) determine genome wide methylated CpG sites by deep sequencing of bisulfite treated genomic DNA to determine sequence context and cytosine methylation variation (BIS-Seq) and RNA (RNA-Seq) and (2) fractionate DNA sequences using differential restriction and/or affinity capture (MMDE-seq) to enrich for methylated DNA sequences. Together, these tools allow targeted interrogation of CpG regions of interest using bioinformatic data mining tools. These technologies will allow creation of a definitive map of the mouse liver methylome from the two parental strains (C57BL/6N and C3H/HeN) and their F1 hybrid (B6C3F1/N) offspring that exhibit dramatically different rates of interstrain and sex dependent spontaneous liver cancers. The high, but variable, incidence of liver tumors in the F1 hybrid mouse often hinders interpretation of 2-year toxicology and carcinogenesis studies. The reference database will aid our understanding of the relationship between variations in sporadic and induced disease incidence associated with individual variations in the methylome, DNA sequence, and exon specific transcript expression critical to understanding the potential functional consequences from generation to generation. With data from this project, the NTP/DIR at NIEHS can create a reference for future investigations into environment-induced changes in methylome variation and its role in spontaneous and induced disease by analysis of archived FFPE tissues with known outcomes and in planning future studies.

This page left intentionally blank

IV.2 Tox21 Activities:

Caenorhabditis elegans "Worm Tox" Screening Facility

This page left intentionally blank

IV.2 *Caenorhabditis elegans* "Worm Tox" Screening Facility

Jonathan H. Freedman, Ph.D.
NIEHS Comparative Genomics Group, Laboratory of Toxicology and Pharmacology, Environmental Toxicology Program

Biomolecular Screening Branch Scientists
Windy Boyd, Ph.D.
Julie Rice
Daniel Snyder
Paul Dunlap

NTP Collaborating Scientists
Mamta V. Behl, Ph.D. (TB)
G. Jean Harry, Ph.D. (NLB)
Michelle J. Hooth, Ph.D. (TB)
Michael Waalkes, Ph.D. (NLB)

NIEHS Collaborating Scientists
Scott Alper, Ph.D. (LRB; currently at Integrated Department of Immunology, National Jewish Health)
David L. Armstrong, Ph.D. (LN)
Grace E. Kissling, Ph.D. (BB)
Lawrence Lazarus, Ph.D. (MCG, LTP)
Richard Paules, Ph.D. (IRG, LTP)
Ben Van Houten, Ph.D. (LMG; currently at the Dept of Med-Pharmacology & Chem Bio, University of Pittsburgh)

Collaborating Scientists
Monica Colaiacovo, Ph.D. (Dept of Genetics, Harvard Medical School, Boston, MA)
Kevin Crofton, Ph.D. (U.S. EPA NHEERL)
Alvin Crumbliss, Ph.D. (Dept of Chemistry, Duke University, Durham, NC)
David Dix, Ph.D. (U.S. EPA NCCT)
Wolfgang Liedtke, M.D., Ph.D. (Center for Translational Neuroscience, Duke University Medical Center, Durham, NC)
Joel Meyer, Ph.D. (CEINT, Duke University, Durham, NC)
William Mundy, Ph.D. (U.S. EPA NHEERL)
Stephanie Padilla, Ph.D. (U.S. EPA NHEERL)
Christopher J. Portier, Ph.D. (NCEH/ATSDR, Atlanta, GA)
Marjolein Smith (SRA International)

IV.2.1 Background and Rationale

There are over 80,000 chemicals in use today and approximately 2,000 new chemicals are introduced into use every year. To assess the potential health effects of such a large number of chemicals,

alternative toxicological methods and models are needed to decrease the time and expense that are associated with current toxicity testing protocols. In 2004, the NTP published a vision statement, *Toxicology in the 21st Century*, outlining the NTP's role in advancing toxicological science by moving away from traditional observational studies to more predictive and mechanistic-based approaches (1). Building on the NTP's vision, a National Research Council (NRC) expert panel released a report from the NAS detailing the panel's recommendations for future toxicity testing. This report recommended that future toxicity testing rely less on traditional toxicological models, such as mammalian species, and move in the direction of pathway-based *in vitro* screening and testing using invertebrate species (2).

IV.2.1.1 History of WormTox

Based on these reports, the NTP, through collaboration with Duke University, began the development of the NTP *C. elegans* Screening Facility (aka WormTox). The initial tasks of WormTox were:

- **Task 1**. Develop methods to measure the toxicity of known and suspected developmental and neurological toxicants in *C. elegans*. This task involved the development of multi-well formatted, high content screening methods and statistical tools for the generation and analysis of growth, reproduction, feeding, and locomotion data.

- **Task 2**. Expose *C. elegans* to at least 200 known or suspected developmental or neurological toxicants and determine quantitative changes in phenotypic characteristics.

- **Task 3**. Create and/or obtain GFP (green fluorescent protein)-based, stress-responsive transgenic *C. elegans* lines for improving sensitivity and specificity of toxicity screens. This task also included the development of multi-dimensional analytical software to quantitatively measure the effects of toxicant exposure on gene expression at the cellular level.

- **Task 4**. Use *C. elegans* microarray analysis and test a subset of chemicals from Task 2 for changes in gene expression. (This task was dropped in the early stage of WormTox development due to cost and personnel limitations.)

- **Task 5**. Adapt methods for medium throughput analysis to assess the toxicological responses in *C. elegans* in which each gene has been inactivated using RNA interference techniques.

In 2005, WormTox moved from Duke University to the Laboratory of Toxicology and Pharmacology in the Division of Intramural Research (DIR) at NIEHS. Following the realignment of the NTP and the formation of the Biomolecular Screening Branch, WormTox moved to its current location within that branch.

IV.2.1.2 C. elegans as a model organism

C. elegans provides an excellent model system for obtaining an integrated picture of cellular, developmental, and molecular aspects of the effects of toxicants on growth and development, as well as gene expression. *C. elegans* are composed of only 959 somatic cells, but they contain highly differentiated muscle, nervous, digestive, and reproductive systems (3,4). The fates and lineages of all somatic cells in embryos, larvae and adults have been established, and the developmental program is invariantly reproduced in each animal (3-6). The nematode genome is relatively small (10^8 bp), completely sequenced, and an abundance of information is available on the genetic and physical maps of its chromosomes (7,8). The genome is thoroughly annotated including GO assignments, and is included in the KEGG, COG, Cytoscape, and Ingenuity databases. Each predicted mRNA is mapped to a

physical location in the genome, and descriptions include similarity scores to a variety of species, including humans.

Several factors combine to make *C. elegans* a practical species for toxicological testing. First, there is a concerted effort by scientists and the public to reduce, replace, or refine the use of mammals in laboratory testing (9,10). Second, as previously described, there is a wealth of knowledge available on *C. elegans* biology. An exceptionally detailed database on the cell and developmental biology including a map of all of the neuronal pathways; the ability to observe by microscopy all of the somatic cells in living *C. elegans;* and the technology to quickly produce transgenic nematodes are available (11-15). In addition, several studies have demonstrated that changes in *C. elegans* following chemical exposure appear to be predictive of developmental shifts and/or neurological damage seen in laboratory studies using rodents (16-27). Finally, new tools in robotics, image acquisition and analysis, gene knockout, and gene and protein expression measurements make it possible to study complex biological processes in a medium-throughput fashion using *C. elegans*. Thus, *C. elegans* can help meet the goals described for the future toxicology: an alternative to mammalian species and a mechanistic approach to toxicological studies.

Several factors of *C. elegans* biology indicate that it can serve as a model in studies of human disease and toxicology. A high degree of evolutionary conservation between *C. elegans* and higher organisms is observed in many signal transduction and stress-response pathways (28-33). For example, homologues of several human proto-oncogenes have been identified in *C. elegans*. Much of our current understanding of the organization of the *ras* signal transduction pathway has been elucidated by forward- and reverse-genetic analyses of *C. elegans* (34). In addition, homologues of tumor suppressor genes involved in renal cell carcinoma, hemangioblastoma and breast cancer have been identified. Many of the regulatory processes controlling apoptosis in higher organisms are conserved, and have been elucidated from studies in *C. elegans*. In fact, the 2002 Nobel Prize in Physiology or Medicine was awarded to three *C. elegans* researchers for their research in this field. Because of the evolutionary conservation of many genes and regulatory pathways, *C. elegans* has provided new information into the mechanisms of human diseases including Menkes and Wilson's diseases, cancer, Alzheimer's disease, neurological disorders, and polycystic kidney-disease (35-48). In addition, homologues for many of the genes induced in response to toxicant exposure in vertebrates have been identified in *C. elegans*. These include metallothionein, superoxide dismutase, ubiquitin, heat shock proteins 16 and 70, glutathione-S-transferase, p-glycoprotein, cytochrome p450 and catalase (28-32,49-53). *C. elegans* also contain homologues to many of the vertebrate signal transduction proteins and pathways that have been implicated in regulating cellular responses to toxicant exposure. Because of the evolutionarily conserved nature of signal transduction and stress-response pathways, it is likely that responses elicited in *C. elegans* will be applicable to understanding similar processes in higher organisms, including humans. It has been suggested that "virtually any gene of interest can be studied at the functional level" in *C. elegans*. In addition, estimates suggest that >30% of the genes in *C. elegans* will have homologues in humans (7,54).

IV.2.1.3 C. elegans and Tox21

WormTox has been a component of the Tox21 initiative since its inception. As part of Tox21, WormTox has several roles: first, as part of a triaging scheme to help identify high priority chemicals that should be tested in higher/mammalian species. WormTox along with other model organisms constitute the link between high throughput *in vitro* screening and testing in traditional mammalian species (55). For example, for 10,000 potential toxicants, *in vitro* screening will identify chemicals that have potential human toxicities; this "short list" will then be tested in WormTox and other non-mammalian species. Chemicals that are defined as toxic using these models will then be tested in rodents. It is a goal of

Tox21 that using this scheme, the 10,000 initial chemicals can be reduced to dozens of potential mammalian toxicants. Second, *C. elegans* and WormTox serve to help define mechanisms of chemical toxicity. As noted above, the wealth of cellular and molecular information on *C. elegans* make it an invaluable tool to define the mechanistic effects of toxicants on growth and development, neurological activity, gene expression, signal transduction, metabolic activities, and physiological functions (feeding, reproduction, etc).

IV.2.2 Key Issue, Hypothesis Tested, or Problem Addressed

The overall problem being address by WormTox is: can *C. elegans* be used as an alternative toxicological test species to help in predicting the toxicity of uncharacterized chemical to humans? To address this problem the major goals of WormTox are:

- Develop multi-well based medium-throughput assays to monitor the effects of toxicants on *C. elegans* phenotypes (growth, feeding, fecundity, locomotion, gene expression).
- Develop robust statistical tools for the analysis of phenotypic data.
- Test a variety of chemicals, individually or as components of chemical libraries.
- Compare the *C. elegans* results to those generated by other Tox21 partners.
- Transmit the results from the WormTox studies to the scientific community and the general public.

The WormTox group has completed several of these goals (see below). Assays to measure growth, feeding and reproduction have been created using 96-well format. A locomotion assay has been developed in 24-well format. For each of these assays, statistical tools have been created, most of which have been generated *de novo*. We have used a set of these assay to test the NTP 1408 and U.S. EPA ToxCast Phase I chemical libraries (56,57). In addition, WormTox has tested individual chemicals for intramural and extramural investigators, and several chemicals that are the focus of current NTP studies. Through a task-order mechanism with Sciome, WormTox is currently developing QSAR models and comparing the data from the *C. elegans* ToxCast Phase I chemical screen with historical rodent data and data generated by other ToxCast partners. Descriptions and applicability of all assays have been presented at national and international meetings, and published in peer-reviewed journals.

IV.2.3 Approach

A critical element in the design and implementation of any testing program is the selection of the biological model to use in identifying a potential hazard. For many years, the NTP has primarily relied upon rodent systems to test for the toxicity of environmental agents. The key component driving this decision has been the close similarity between rodents and humans in terms of the critical components governing toxicity such as the physiology, biochemistry, and overall structure of the biological systems. Numerous papers have been written comparing these systems, and scientific and regulatory methods have been established to predict response in humans from responses observed in rodents.

The recent development of better molecular and computational tools has provided an opportunity for the development of faster screens in genetically modified rodents or in other species. The ability to develop targeted organisms for testing hypotheses is well established in the scientific literature, and has led to substantial insight into the causal mechanisms, therapy, and prevention of disease. Much of this

work has been done by independent researchers interested in a particular hypothesis or mechanism they wish to address. There has been very little effort aimed at the translation of these basic research tools into the types of tools necessary for a testing program of the magnitude of the NTP. To begin translating basic toxicological research tools to a more applied testing program, WormTox uses the free-living soil nematode *C. elegans* as an invertebrate, toxicological model to assess the effects of potential developmental and neurological toxicants on multi-cellular organisms.

A number of phenotypes have been observed in *C. elegans* after chemical exposures including changes in behavior, morphology, and gene expression. Based on these phenotypes, we determined endpoints of interest and the availability of appropriate technologies to increase the throughput and automation of measurements of *C. elegans* phenotypes for the development of rapid, low-cost assays to measure the toxicity of known and suspected toxicants.

After development of each assay, *C. elegans* are exposed to toxicants with known effects on nematode phenotypes (survival, size, growth, reproduction, and locomotion) to verify that these phenotypes can be reliably measured in semi-automated, medium-throughput format. Finally, chemical libraries with dozens to hundreds of chemicals are used to determine appropriate experimental design and statistical analyses for high throughput screens.

To measure toxicant-associated phenotypic changes in *C. elegans* in a rapid fashion, the COPAS Biosort flow sorting system is used. The COPAS can make observations of up to 6,000 *C. elegans*/per minute and is designed to sort, dispense, and measure various parameters of individual nematodes (58). In the Biosort, nematodes pass through a flow cell where up to four attributes for each individual may be assessed: time of flight (TOF), which relates to nematode length; extinction (EXT), which corresponds to the optical density; and two fluorescence measurements **(Figure IV.2-1)**. Both TOF and EXT measurements are related to the age and size of the nematode; and both increase as *C. elegans* develop through larval stages into adults (58). **Figure IV.2-1** illustrates the growth of untreated, wild-type *C. elegans* sampled at 24-hour (h) intervals as measured by a COPAS Biosort. The COPAS is used in the feeding, reproduction, and growth assays.

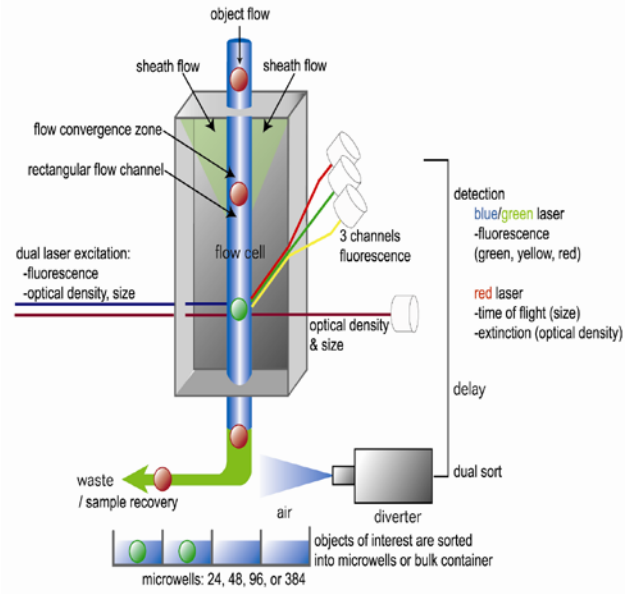


Figure IV.2-1 Diagram of the object flow path and sorting technique of the COPAS Biosort

IV.2.4 Significance

We believe that the information and concepts obtained from the work performed by the NTP *C. elegans* Screening Facility will expand the area of mechanistic toxicology. It will directly address the underlying mechanisms responsible for environmentally induced diseases that are associated with exposure to toxicants. Specifically, the studies performed by WormTox will make significant contributions to the efforts of the Biomolecular Screening Branch and Tox21 partners to help transition toxicology from an observational science to a more mechanism-based science. Through the integration of high throughput

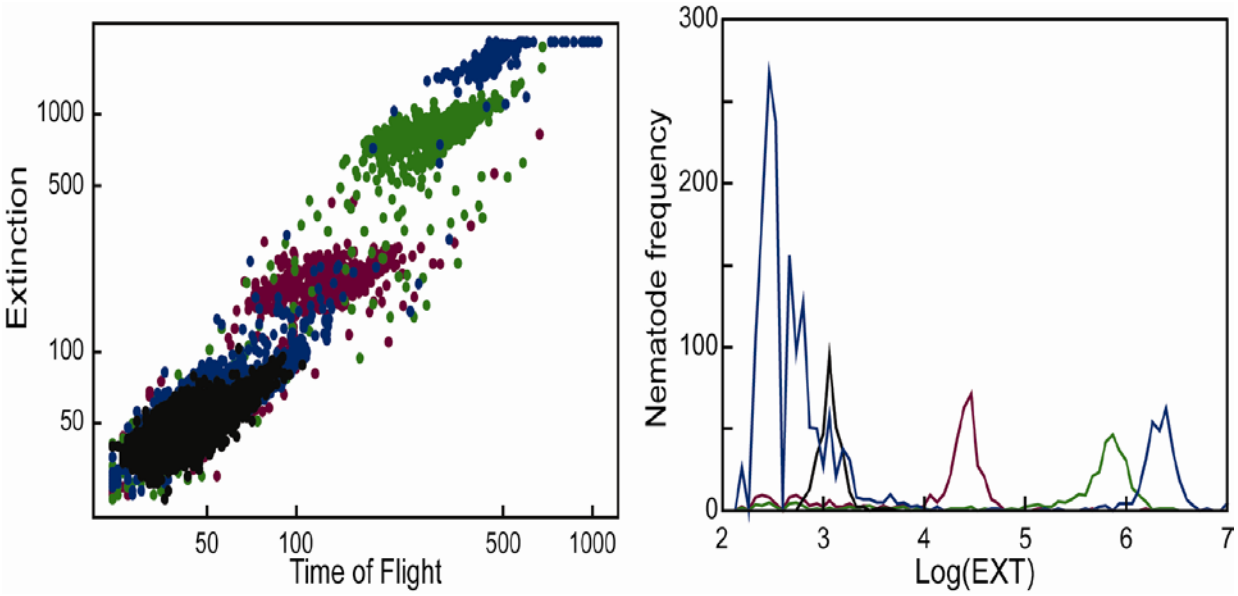


Figure IV.2-2 Untreated L1 nematodes sampled at 0 (black), 24 (red), 48 (green), and 72 h (blue). At 72 h, adult nematodes (high EXT and TOF) and their offspring (low EXT and TOF) were observed. *left panel*, each point corresponds to an individual nematode. *right panel*, frequency distributions of log(EXT) versus numbers of nematodes.

in vitro data with the results obtained from *in vivo* alternative species studies, mechanism of toxicant activity can be defined. This will ultimately be integrated into predictive toxicological models.

In addition to its significance in the area of mechanistic toxicology, WormTox will add to the continuing effort of the NTP, and other national and international agencies to reduce, refine, and replace the number of mammalian species needed for toxicological testing.

IV.2.5 Results/Progress

The WormTox group has developed four *C. elegans* toxicological assays, which readily lend themselves to mechanistic studies of toxicant actions: feeding, growth, reproduction, and locomotion. Each assay was designed to measure the effects of chemicals on a specific phenotype at the appropriate developmental stage. The phenotypes used to develop WormTox assays are regulated by various endogenous and external factors such as temperature and food availability, and have been previously characterized using direct observation by microscopy. The original methods are useful for detailed studies of a few genes or chemicals at a time, but are time-consuming and often subjective. WormTox assays are designed to collect quantitative measurements of *C. elegans* phenotypes while also increasing reliability and throughput. All of the assays were validated by direct microscopic observation. This confirmed that results obtained by high content analysis correlated with the observed phenotypes.

IV.2.5.1 C. elegans feeding assay

C. elegans feeding results from the coordinated pumping of two bulbs of the pharynx (**Figure IV.2-3**), pushing bacteria suspended in liquid into the intestine while simultaneously ejecting excess liquid anteriorly (59). In the WormTox feeding assay, adult *C. elegans* are exposed to chemicals, followed by the addition of red fluorescent microspheres. After ingestion of the microspheres, the red fluorescence is quantified in individual nematodes by the COPAS Biosort. As illustrated with chlorpyrifos oxon, feeding as measured by fluorescence decreases with exposure to increasing concentrations of neurotoxicants (**Figure IV.2-3**). The *C. elegans* feeding assay has been published in *PLoS One* (60).

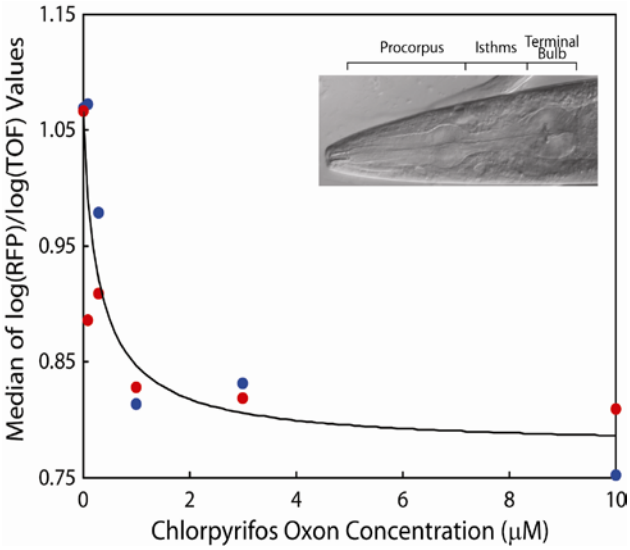


Figure IV.2-3 Effects of chlorpyrifos oxon on *C. elegans* feeding - For each of the two replicate experiments (shown as blue and red), groups of 25 adult *C. elegans* were exposed to one of five concentrations of chlorpyrifos oxon (plus controls) for 24 h. Medians of log(RFP)/log(TOF) values are plotted and the data fit to the Hill equation. *inset*, adult *C. elegans* pharynx.

IV.2.5.2 C. elegans growth assay

Development of *C. elegans* proceeds from the fertilized embryo through four larval stages (L1-L4) to the gravid adult hermaphrodite in approximately 3 days at 20°C. Each of the larval stages can be identified by size and the development of distinct morphological features, such as the pharynx and gonad. The WormTox growth assay uses the COPAS Biosort to dispense L1 nematodes and then measure their size after 48-h exposures, at which time untreated nematodes have developed to the L4 or last larval stage.

Figure IV.2-4 illustrates the effects of chlorpyrifos on *C. elegans* growth. The *C. elegans* growth assay has been published as a pair of manuscripts in PLoS One (61, 62).

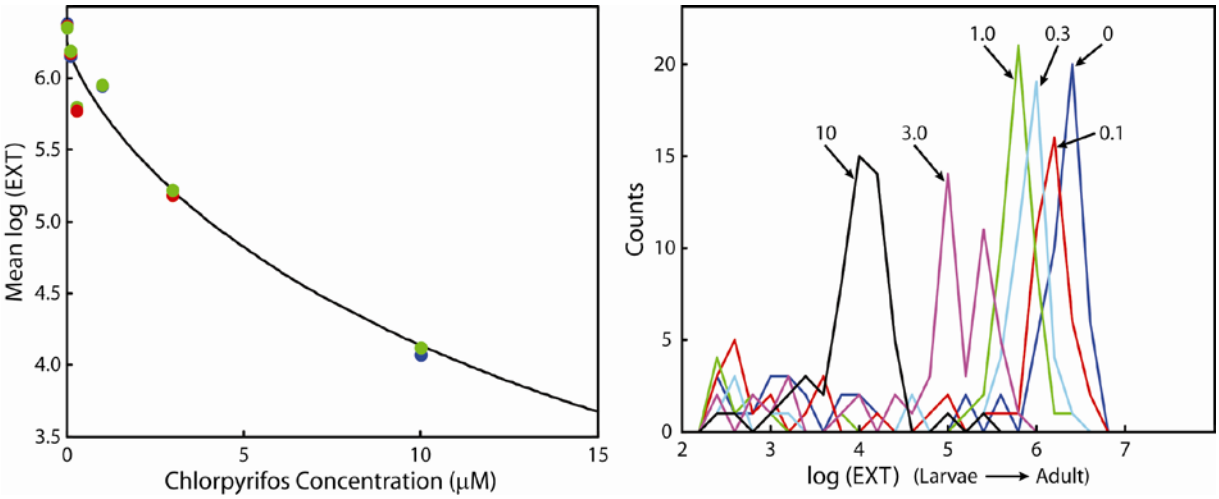


Figure IV.2-4 Effects of chlorpyrifos on *C. elegans* growth—For each of three replicate experiments (shown as blue, red, and green), groups of 50 L1 nematodes were exposed to one of five concentrations of chlorpyrifos for 48 h. *left panel*, observations measured and estimated means fit to Hill equation. *right panel*, modeled frequency histograms showing the effects of different concentrations of chlorpyrifos (0, 0.1, 0.3, 1.0, 3.0, and 10 µM) on nematode size distribution (log(EXT)) for one of the replicates.

IV.2.5.3 C. elegans reproduction assay

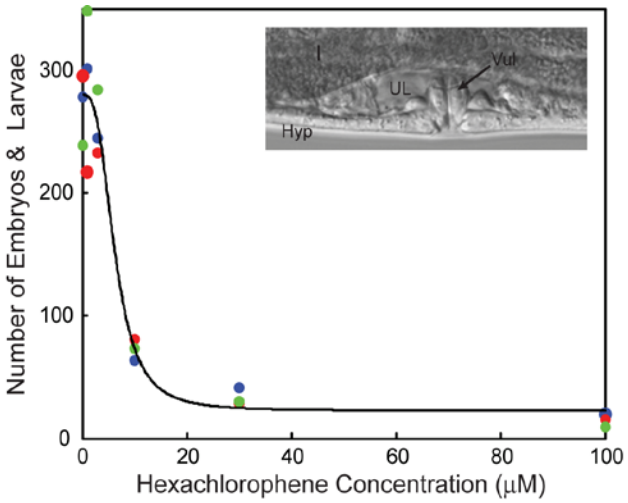


Figure IV.2-5 Effects of hexachlorophene on *C. elegans* reproduction - For each of three replicate experiments (shown as blue, red, and green), groups of five L4 nematodes were exposed to hexachlorophene for 48 h. The numbers of observations (larvae and embryos) were fit to the Hill equation. *inset*, region around a young adult *C. elegans* vulva: I, intestine; UL, uterine lumen; Hyp, hypodermis; Vul, vulva.

C. elegans hermaphrodites first produce sperm as L4s and then oocytes as adults, which eventually form the fertilized embryo throughout the first several days of adulthood. After fertilization, developing embryos are expelled by contraction of the vulval muscles (**Figure IV.2-3, inset**). In the WormTox reproduction assay, nematodes are exposed for 48 h starting at the L4 stage. The total number of offspring and adults are then counted using the COPAS Biosort. The effect of hexachlorophene on *C. elegans* reproduction is presented in **Figure IV.2-5**. The *C. elegans* reproduction assay has been published in *Toxicology and Applied Pharmacology* (63).

IV.2.5.4 C. elegans locomotion assay

A locomotion assay has recently been developed to quantify *C. elegans* movement using a semi-automated motion tracking system. This system consists of an inverted fluorescence microscope equipped with a CCD camera, incubated motorized stage, and image analysis software. Over twenty locomotion parameters have been defined including

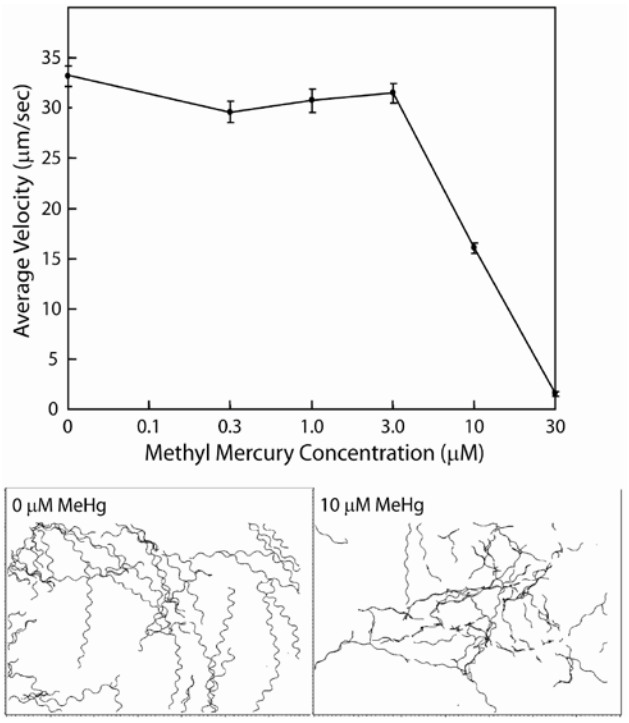


Figure IV.2-6 Effects of methyl mercury on *C. elegans* movement - L4 nematodes were tracked after exposure to methyl mercury for 4 h. *upper panel*, average velocity of *C. elegans* motion. *lower panels*, computer-generated tracks for control nematodes (*LEFT*) and exposed (*right*).

curvilinear velocity, linearity, and amplitude of sinusoidal movement. The locomotion assay has been used to test the effects of several metals and pesticides, as well as a number of loss-of-function phenotypes after RNA interference of genes known to result in abnormal movement. **Figure IV.2-6** illustrates the concentration-dependent effects of methyl mercury. A manuscript is currently in preparation describing the development and application of the *C. elegans* locomotion assay.

IV.2.5.5 Chemicals tested

To date, the WormTox group has tested the effects of over 65 compounds on *C. elegans* growth, reproduction, and feeding including transition metals, solvents, DNA damaging agents, oxidizing agents, and a suite of organophosphate pesticides (**Table IV.2-1**). A number of compounds that were nominated for testing in NTP rodent bioassays have also been examined by WormTox. Working with NTP toxicologists during the study design phase, the appropriate experimental designs for *C. elegans* assays are identified including the number of compounds, concentrations, and *C. elegans* assays to test. For example, *C. elegans* reproduction, growth, and feeding assays were performed to compare the effects of four ionic liquids nominated to the NTP: 1-ethyl-3-methylimidazolium chloride, 1-butyl-3-methylimidazolium chloride, 1-butyl-1-methylpyrrolidinium chloride, and N-butylpyridinium chloride. WormTox also compared the effects of four pyridine compounds and will soon test a number of flame retardants, all of which will also be tested in rodent bioassays. Once the results from NTP studies are completed, the results from rodent studies will be compared with those from WormTox assays to assess how the *C. elegans* can best complement traditional NTP rodent bioassays.

Table IV.2-1

Chemicals tested in medium-throughput growth, reproduction, and feeding assays

Metals	Organics		
Aluminum Chloride	1-Butyl-1-methyl Pyrrolidinium Cl	Demeton-S-Methylsulfone	Methylisothiocyanate
Cadmium Chloride	1-Butyl-3-Methylimidazolium Cl	Dichlorvos	MMS
Chromium Oxide	1-Ethyl-3-Methylimidazolium Cl	Diphenylhydantoin	MNNG
Cobalt Chloride	2-Aminopyridine	Diquat	Monocrotophos
Copper Sulfate	3-Aminopyridine	DMSO	N-Butylpyridinium Chloride
Lead Acetate	4-Aminopyridine	EMS	Nicotine
Lead Nitrate	Acetaminophen	ENU	Paraquat
Manganese Oxide	Acetic Acid	Ethephon	Parathion
Mercuric Chloride	All-trans Retinoic Acid	EtOH	PCB mixture
Methyl Mercury	α-Cyclodextrin	Fumonisin	PEG-60
Nickel Sulfate	Ascorbic Acid	Glyphosate	Pyridine
Silver Nitrate	β-Cyclodextrin	Juglone	Sodium Metam
Sodium Arsenite	β-Cyclodextrin Hydrate	Lindane	Tamoxifen
Sodium Selenite	Bisphenol-A	Methadone	Tebuconazol
Thimerosol	Caffeine	Methanol	Valproic acid
Vanadium Oxide	Carbaryl	Methyl Cellulose	
Zinc Sulfate	Chlorpyrifos	Methyl Parathion	

IV.2.5.6 Analysis of chemical libraries

For chemical libraries containing hundreds or thousands of chemicals, the WormTox Group uses medium-throughput *C. elegans* assays to evaluate chemicals' toxicological activities. To accommodate the large number of chemicals tested, the reproduction and growth assays were modified to rapidly screen compounds. Using the modified reproduction assay, the WormTox group screened the effects of a chemical library containing 1408 compounds selected by the NTP (57); these compounds were classified as active, inactive, or inconclusive based on their effects relative to untreated groups.

In a recent screen of the U.S. EPA's ToxCast Phase I chemical library (64), a modified *C. elegans* growth assay was applied to determine toxicity values depending on the severity of chemical effects. The library, which consists mainly of pesticide actives, was tested across seven concentrations from 0.5 - 200 µM. Chemical activity was described using the slope of the concentration-response plot as well as the decrease in size relative to untreated controls. Therefore, the chemicals with the most negative slope and smallest size were the most active chemicals, while inactive chemicals resulted in no slope or larger size. Almost 65% of the chemicals caused some effects on *C. elegans* growth with approximately half

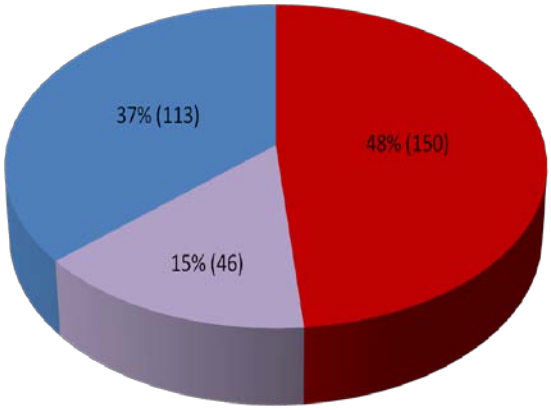


Figure IV.2-7 Activity of 309 ToxCast compounds in *C. elegans* growth assay. Red, active; blue, inactive; purple, inconclusive

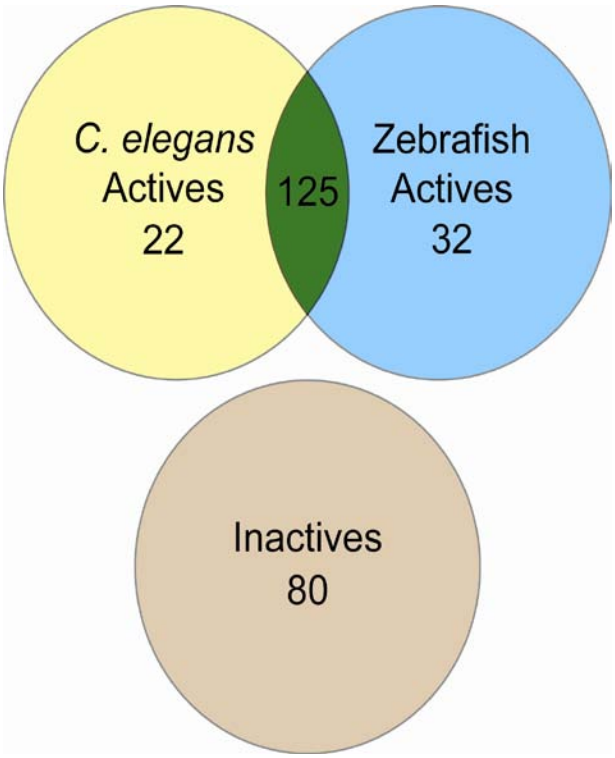


Figure IV.2-8 Concordance between *C. elegans* and zebrafish *in vivo* toxicity results. Matthews correlation coefficient = 0.5759.

causing severe decreases such that lethality or no growth was observed after exposures (**Figure IV.2-7**). In the case of monotonic decreases in growth, EC_{50}s were calculated; otherwise, lowest effective levels (LEC) and percent decreases in growth were calculated. The results from the screening of the ToxCast Phase I chemical library are currently being prepared for publication.

IV.2.5.7 Comparisons of C. elegans data to other species

The ToxCast Phase I library is composed of well-characterized chemicals with a large amount of toxicity data and has now been screened in hundreds of *in vitro*, biochemical, and alternative animal model assays. To date, two preliminary comparisons have been made between the *C. elegans* data and other assays: zebrafish activity data and rodent database values.

❖ **Zebrafish**

First, in collaboration with the U.S. EPA, chemical activity was compared between *C. elegans* growth and zebrafish embryonic development. Over 40% of the compounds caused developmental defects in zebrafish and decreased growth in *C. elegans,* while 30% were not toxic to either at the maximum concentrations tested (**Figure IV.2-8**). The class of pesticides with the highest number of toxic compounds included conazoles, pyrethroids, and organophosphates. The compounds' activities in both organisms were highly correlated with log octanol-water partition coefficients and molecular weights.

❖ **Rodent**

The *C. elegans* growth data was also compared to a number of mammalian endpoints in the U.S. EPA's Toxicity Reference Database (ToxRefDB), developed by the U.S. EPA NCCT in partnership with U.S. EPA's Office of Pesticide Programs. The *in vivo* toxicity endpoints included in ToxRefDB are based on chronic, sub-chronic, cancer, developmental and reproductive toxicity experiments. Chemicals with a LEL listed for at least one of six rodent endpoints were included in the comparison. Chemicals with the highest activity in the *C. elegans* growth assay led to the greatest reductions in size, while chemicals with the highest activity in the rodent tests had the

lowest LELs. Therefore, as expected, the correlations between the C. elegans activity score and rodent LELs were negative (**Table IV.1-2**).

In spite of the discrete nature of the measures compared, these results show a very significant consistency between rodent and nematode response to the chemicals. The two rodent endpoints with the weakest correlation (Rabbit dDEV and mouse cancer) were also the two that had the fewest chemicals. Generally, the most active chemicals in the *C. elegans* growth assay corresponded to very low rodent LELs.

Table IV.2-2

Comparison between *C. elegans* and rodent Toxicity data

Rodent end point	Number of chemicals	Kendal's Tau	p-value
Rat cancer	188	-0.26	1.3e-6
Mouse cancer	139	-0.034	0.58
Rat dDEV	142	-0.2	0.0013
Rat mDEV	218	-0.22	1.4e-5
Rabbit dDEV	102	-0.19	0.01
Rabbit mDEV	212	-0.18	0.00052

dDEV = fetal effects (weight reduction, defects); mDEV = maternal effects (weight gain, pregnancy loss)

IV.2.5.8 Statistical and mathematical tools

After creation of monitoring protocols, statistical analysis routines had to be generated specifically for *C. elegans* COPAS and motion tracking data. Because of the large volumes of data collected in WormTox studies, as well as unique characteristics of the machine measurements, we have collaborated with statisticians and mathematicians to develop models to characterize the effects of chemicals on *C. elegans* growth, reproduction, feeding, and locomotion.

Generally, COPAS measurements are plotted against concentrations of the compound tested, and these dose-response curves are used to calculate EC_{50} values, if possible. For the growth assay, the EXT of nematodes as measured by the COPAS is used as a measure of nematode size. A mixed distribution model is applied consisting of a lognormal distribution for noise and a Markov population growth model for nematode measurements.

IV.2.5.9 Extra-and intramural collaborations

Our collaborations have required modification of standardized WormTox protocols to investigate a number of basic research questions outside of toxicological screening. With Drs. Bennett Van Houten's and Joel Meyer's laboratories at NIEHS, *C. elegans* was used as a model system to study DNA damage and repair after exposure to environmental stressors. Along with qPCRs used to quantify DNA damage, the effects of UV-C on *C. elegans* growth, lifespan, and morphology were assessed by WormTox. In collaboration with Dr. David Armstrong's group, the *C. elegans* pharynx was used as a model system to study potassium channel regulation in order to better understand mammalian cardiac functioning.

Other collaborations included Dr. Alvin Crumblis, Department of Chemistry, Duke University, on the ability of metal chelators to affect the toxicity of transition metals in *C. elegans*; and Dr. Monica Colaiacovo, Harvard Medical School, on the effect of toxicants on chromosome segregation.

IV.2.6 Future Directions/Plans and Justifications

There are three major future plans for WormTox: First, a comparison between the *C. elegans* growth data to *in vivo* and *in vitro* data for the chemicals in the ToxCast Phase I library; second, the generation of a collection of ~40 chemical/stress responsive transgenic strains of *C. elegans*; and third, posting the *C. elegans* toxicology data in publicly accessible databases.

IV.2.6.1 Comparison of C. elegans data to other in vivo and in vitro ToxCast Phase I datasets

Comparing *C. elegans* results with other systems is the critical next step in the evaluation of the usefulness of *C. elegans* and other alternative species in the prioritization of further chemical testing. The information generated through this collaboration will be applicable to other ToxCast and Tox21 efforts including comparisons among *in vitro* and mammalian data.

Proposed activities - The need to define the relationship between the results obtained from any screening effort to human health and safety still remains. The U. S. EPA's ToxCast program is investigating this relationship by screening in Phase I a defined set of three hundred compounds, which currently have whole animal toxicity data, using *in vitro* cell and biochemical systems, and alternative whole organism models (56). Although data generation is rapid, the ability to compare *in vitro* and alternative species data to available mammalian toxicity data requires a significant effort. To accomplish this goal, the creation of new and novel approaches in data analysis, mathematical modeling, and software development will be required. The goal of this plan is to develop models and software for the analysis of the *C. elegans* toxicological data. To this end, a QSAR workflow, primarily developed at the Laboratory for Molecular Modeling at the University of North Carolina at Chapel Hill (UNC-CH), to build computational models for the ToxCast Phase I compounds will be employed. Compounds will be characterized by both conventional chemical descriptors calculated from chemical structure, as well as by measured *C. elegans* toxicity end-points, which will be employed as biological descriptors. Procedures established in a recent study of chemical structure – short term assay – long term toxic effect relationship modeling will be followed (65). The results of this study will establish both the absolute, as well as the relative (i.e., with respect to ~600 other short term, *in vitro* assays included in the ToxCast Phase I database) value of *C. elegans* toxicity for predicting chemical *in vivo* toxicity and prioritizing new chemicals for *in vivo* studies.

Relevant animal toxicity data will be obtained from ToxRefDB. Each of the curated set of Phase I compounds has experimental results of up to 78 various *in vivo* toxicity endpoints in ToxRefDB. Most of the *in vivo* endpoints have low ratios of active to inactive compounds, making them unsuitable for QSAR modeling. Thus, 18 out of 78 *in vivo* endpoints, which had the highest active response ratio, will be used. In the original ToxRefDB record, the toxicity is reported as LEL values in the units of mg/kg/day. However, for this analysis, the U.S. EPA NCCT's definitions of compounds as either active (toxic) or inactive (non-toxic) will be used. The number of ToxCast Phase I compounds in each toxicity endpoint subset ranges from 224 to 235, and the active compound fraction ranges from 17.4% to 44.6%.

QSAR models, using the methods of classification such as Random Forrest, SVM with linear kernel and SVM with radial basis function (RBF) kernel will be developed. To ensure the robustness and reliability of the resulting model, 5-fold cross validation will be employed. Earlier studies have demonstrated that the use of biological descriptors of chemicals typically affords QSAR models with higher external predictivity than models developed with conventional chemical descriptors. Thus, the models will be developed for all 18 *in vivo* end points using the following compound parameters: (a) chemical descriptors only; (b) *C.*

elegans toxicity end point only; (c) hybrid chemical/biological representation or the short- term assay – *in vivo* toxicity relationship. The predictive power of models generated under scheme (c) that utilize *C. elegans* toxicity data versus those generated with other short terms assays as a matter of establishing the relative value of the *C. elegans* toxicity assays as part of ToxCast program will be compared.

This work began in September 2010 through collaboration between investigators at the UNC-CH, mathematicians/statisticians at Sciome, LLC, and contract support at SRA International. The tasks to be completed as part of this collaboration include:

- **Task 1**. Curate the ToxCast Phase I chemical library to select unique organic molecules for chemical descriptor calculations.
- **Task 2**. Calculate several sets of chemical descriptors.
- **Task 3.** Develop QSAR models of *C. elegans* end points for the curated ToxCast Phase I dataset.
- **Task 4**. Develop QSAR models of 18 ToxRefDB *in vivo* toxicity end points for the curated ToxCast Phase I dataset.
- **Task 5**. Evaluate the predictive power of *C. elegans* toxicity end points in comparison with ToxCast Phase I *in vitro* assays with respect to selected ToxRefDB *in vivo* endpoints.
- **Task 6**. Develop QSAR models of selected ToxRefDB *in vivo* end points using *C. elegans* toxicity end point data. (A) develop hybrid QSAR utilizing chemical descriptors in combination with *C. elegans* end point data only or with *C. elegans* and other *in vitro* end point data treated as biological descriptors; (B) develop hierarchical two step QSAR models using the relationship between *C. elegans* end points and ToxRefDB *in vivo* toxicities.
- **Task 7**. Use resulting models to prioritize Phase II compounds for both *C. elegans* as well as animal *in vivo* toxicity studies.

Current Progress - The U.S. EPA ToxCast Phase I program provided data for 320 substances (309 unique) with known *in vivo* toxicity measured in 76 assays; the results of ~600 *in vitro* assays for the same substances were available as well. The latter data have been used previously with varying degree of success to improve the predictive power of *in silico* models of chemical toxicity. We have explored, in the same context, new whole organism toxicity data generated for ToxCast Phase I chemicals in the *C. elegans* growth assay. The goal of this study was to establish both the absolute as well as relative (i.e., with respect to other short term assays included in the ToxCast Phase I database) value of *C. elegans* growth assay for predicting chemical *in vivo* toxicity and prioritizing new chemicals for *in vivo* studies. Unlike most *in vitro* assays where the number of active compounds was typically lower than that of inactive ones, the *C. elegans* assay resulted in nearly balanced dataset. kNN (k nearest neighbor) QSAR modeling of *C. elegans* data using standard computational workflow with the emphasis on external validation resulted in models with the total accuracy of 66%. The *C. elegans* data were further explored in the following novel hierarchical QSAR modeling workflow. First, all chemicals were partitioned into two classes based on whether a compound tested similarly (Class I) or dissimilarly (Class 2) in both *C. elegans* and three mouse *in vivo* assays resulting in 96, 109, and 117 of Class I and 117, 112, and 100 Class II compounds. Second, each *in vivo* end point classification models to distinguish toxic versus non-toxic compounds within each Class were developed using kNN binary QSAR approaches and the total external accuracy was in the range of 68-71%. This exceeded that of the same models generated with chemical descriptors only or using other ToxCast Phase I *in vitro* data.

IV.2.6.2 Generation of chemical/stress responsive transgenic C. elegans

The *in vivo* expression of fluorescent reporter proteins, such as GFP, mCherry, and dTomato under the control of endogenous promoters is frequently used to study spatial and temporal gene expression patterns in transgenic *C. elegans*. The fluorescent expression patterns of individual nematodes can be observed via microscopy and rapidly quantified using COPAS flow cytometry. A collection of stable, integrated strains of transgenic *C. elegans* are being generated to monitor the effects of toxicants on gene expression *in vivo*. This collection includes a representative set of target genes from signaling pathways known to be affected by exposure to environmental agents (e.g., DNA damage response, unfolded protein response, apoptosis, receptor mediated signaling). The transgenic *C. elegans* will be used for rapid gene expression profiling to identify signaling pathways and molecular events that are perturbed following contaminant exposure.

Proposed activities – Through the NIH SBIR program, a contract has been awarded to Knudra Tech (Salt Lake City, UT) to create 34 transgenic strains of *C. elegans* that express unique stress-responsive genes. There will be three independent lines for each gene, yielding a total of 102 transgenic *C. elegans*. Each transgenic strain will express dual-reporter genes. The first reporter will serve as a control to assess the effects of chemicals on fluorescence and as an internal standard, which can be compared to the second reporter. The second reporter will consist of the stress/chemical responsive genes. For technical reasons the control reporter will contain *unc-47*::GFP. *unc-47* encodes the *C. elegans* vesicular GABA transporter and is constitutively expressed in neurons (**Figure IV.2-9**). The responsive genes will drive the expression of mCherry (RFP).

We are interacting with Knudra on the selection of genes that will be used as reporters. Gene selection will be made based on conserved pathways outlined in the NAS report entitled *Scientific Frontiers in Developmental Toxicology and Risk Assessment* (66). In addition, input is being solicited from other members of the Biomolecular Screening Branch and the Tox21 partners in order to potentially evaluate the same pathways in qHTS at the NCGC and in *C. elegans*.

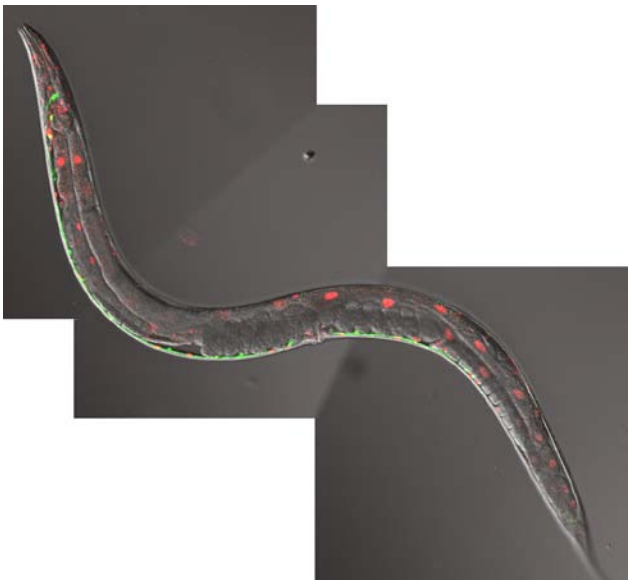


Figure IV.2-9 Heat-shocked transgenic adult *C. elegans* expressing phsp-16.41::mCherry and punc-47::GFP.

One of the major challenges in this project will be the rapid acquisition and analysis of fluorescence data from transgenic *C. elegans*. Data acquisition will be accomplished using a recently-obtained add-on to the COPAS Biosort, the Profiler, which allows for the quantification of fluorescence in optical slices collected along the length of an individual nematode. Recently, the bioinformatics/statistical consulting contract for the NTP and NIEHS has been awarded to SRA International. We have been meeting regularly with SRA staff and are discussing potential approaches by which the fluorescence data could be analyzed. Based on our past successes with SRA, we anticipate that we will successfully develop the tools necessary to quantitatively assess the effects of toxicant exposure on transgene expression.

Table IV.2-3
Transgenes being developed at Knudra

C. elegans Gene	Common Name	Function
skn-1	NRF2/	regulates oxidative stress response
gcs-1	glutamine cysteine synthetase	glutathione bioshyntesis
ugt-1	UDP-glucuronosyl transferase	Phase 2 metabolism
gst-38	glutathione-S-transferase	Phase 2 metabolism
hsp-4	endoplasmic reticulum chaperone	ER stress response
hsp-60	mitochondrial-specific chaperone	mitochondrial stress response

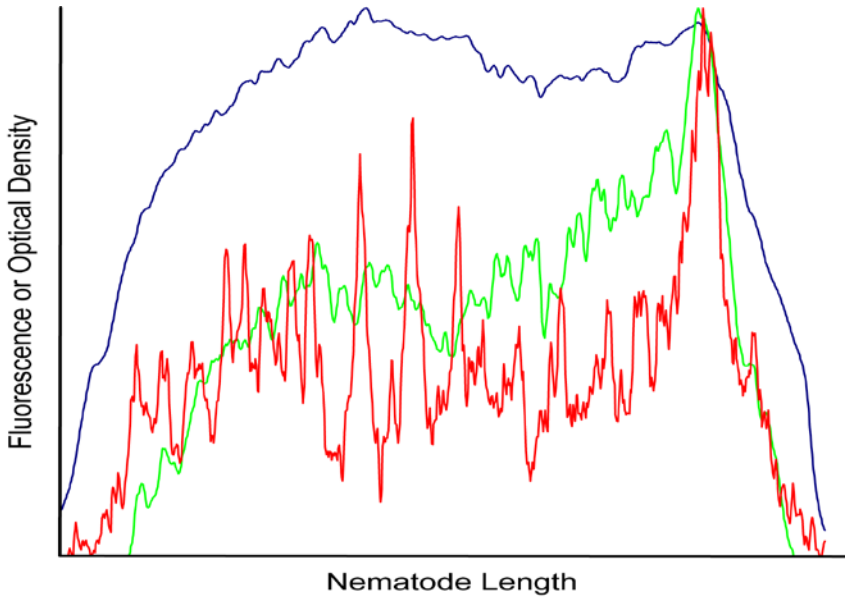


Figure IV.2-10 Profile of one heat-shocked Phsp-16.41::mCherry nematode. Optical density (also called extinction) is measured across the length (or time of flight) of individual animals (Blue). Fluorescence is detected in 2 channels at 514 nm. red, mCherry; green, GFP fluorescence.

Current Progress –The Knudra contract began July 2010. We have sent a list of the first six genes to be used in generating transgenic *C. elegans* (**Table IV.2-3**).

We have received a transgenic strain containing the low molecular weight heat shock protein gene, *hsp-16*, which controls the expression of mCherry (RFP). This strain will be used to design future experiments using the COPAS. Preliminary tests with this strain showed a robust response to heat shock (**Figure IV.2-9**). In addition, COPAS Profiler data (**Figure IV.2-10)** was obtained indicating that this strain can be used in the initial development of experimental designs and statistical tools.

IV.2.6.3 Posting C. elegans toxicology data to public databases.

A future goal is to make the WormTox data available to other members of the Tox21 community and external researchers through development of a publicly accessible database. Within each experiment, multiple data files are generated that include COPAS Biosort data files, experimental notes on design, organism health, chemical preparation, statistical analysis, and graphical summaries. Currently, experimental notes from archived data are being converted to digital records. As the results of experiments are published, we plan to make all data and statistical tools available to the public. The

WormTox data can then be compared to other toxicological models and used to inform future toxicological screens.

Through collaboration with Dr. Jennifer Fostel (NTP), we plan to develop a *C. elegans* database within the CEBS database. To start, we will modify the existing CEBS interface to include WormTox-specific assay conditions and data files. Example data files from each type of COPAS assay (growth, feeding, and reproduction) along with assay descriptions will be provided to the CEBS support group. Once the WormTox interface has been developed, the *C. elegans* growth data from the ToxCast Phase I screen will be provided to CEBS and publicly-released after publication.

IV.2.7 Publications

The WormTox group has in press five primary and made significant contributions to seven additional peer-reviewed publications. WormTox has made over 50 presentations at national and international meetings.

Boyd WA, McBride SJ, Freedman JH. 2007. Effects of genetic mutations and chemical exposures on *Caenorhabditis elegans* feeding: Evaluation of a novel, high-throughput screening assay. PLoS One 12, e1259. (PMID: 18060055)

Alper S, Laws R, Lackford B, Boyd WA, Dunlap P, Freedman JH, et al. 2008. Identification of innate immunity genes and pathways using a comparative genomics approach. Proc Natl Acad Sci U S A 105:7016-7021. (PMID: 18463287)

Peterson RT, Nass R, Boyd WA, Freedman JH, Dong K, Narahashi T. 2008. Use of non-mammalian alternative models for neurotoxicological study. Neurotoxicology 29:546-555. (PMID: 18538410)

Smith MV, Boyd WA, Kissling GE, Rice JR, Snyder DW, Portier CJ, et al. 2009. A discrete time model for the analysis of medium-throughput *C. elegans* growth data. PLoS One 4, e7018. (PMID: 19753303)

Boyd WA, Smith MV, Kissling GE, Rice JR, Snyder DW, Portier CJ, et al. 2009. Application of a mathematical model to describe the effects of chlorpyrifos on *Caenorhabditis elegans* development. PLoS One 4, e7024. (PMID: 19753116)

Boyd WA, Crocker TL, Rodriguez AM, Leung MC, Lehmann DW, Freedman JH, et al. 2010. Nucleotide excision repair genes are expressed at low levels and are not detectably inducible in *Caenorhabditis elegans* somatic tissues, but their function is required for normal adult life after UVC exposure. Mutat Res 683:57-67. (PMID: 19879883)

Boyd WA, McBride SJ, Rice JR, Snyder DW, Freedman JH. 2010. A high-throughput method for assessing chemical toxicity using a *Caenorhabditis elegans* reproduction assay. Toxicol Sci 245:153-159. (PMID: 20206647)

Boyd WA, Freedman JH. 2010. Medium- and high-throughput toxicity screens using *C. elegans*. ALTEX 27(special issue):79-85.

Leung MCK, Goldstone JV, Boyd WA, Freedman JH, Meyer JN. 2010. *Caenorhabditis elegans* generates biologically relevant levels of genotoxic metabolites from aflatoxin B_1 but not benzo(a)pyrene in vivo. Toxicol Sci; doi: 10.1093/toxsci/kfq295 [Online 23 September 2010]. (PMID: 20864627)

Boyd WA, Smith MV, Kissling GE, Freedman JH. 2010. Medium- and high-throughput screening of neurotoxicants using *C. elegans*. Neurotoxicol Teratol 32:68-73. (PMID: 19166924)

IV.2.8 References

1. National Toxicology Program. 2005. Toxicology in the 21st Century: The Role of the National Toxicology Program. http://ntp.niehs.nih.gov/ntp/main_pages/NTPVision.pdf

2. National Research Council (U.S.). Committee on Toxicity Testing and Assessment of Environmental Agents. 2007. Toxicity testing in the 21st century : a vision and a strategy. National Academies Press, Washington, DC.

3. Sulston J. 1988. Cell Lineage. In: The Nematode *Caenorhabditis elegans* (Wood WB, ed). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press,, 123-155.

4. Kenyon C. 1988. The nematode *Caenorhabditis elegans*. Science 240(4858):1448-1453.

5. Sulston J, Horvitz RH, Kimble J. 1988. Cell Lineage (Appendix). In: The Nematode *Caenorhabditis elegans* (Wood WB, ed). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 457-489.

6. Lambie EJ. 2002. Cell proliferation and growth in *C. elegans*. Bioessays 24(1):38-53.

7. The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 2825396): 2012-2028.

8. Harris TW, Chen N, Cunningham F, Tello-Ruiz M, Antoshechkin I, Bastiani C, et al. 2004. WormBase: a multi-species resource for nematode biology and genomics. Nucleic Acids Res 32 (Database issue):D411-417.

9. Russell WM, Burch RI. 1959. The Principles of Human Experimental Technique. Methuen. London.

10. Mehlman MA, Pfitzer EA, Scala RA. 1989. A report on methods to reduce, refine and replace animal testing in industrial toxicology laboratories. Cell Biol Toxicol 5(3):349-358.

11. Waterston R, Sulston J. 1995. The genome of *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 92(24):10836-10840.

12. Mello C, Fire A. 1995. DNA transformation. Methods Cell Biol 48:451-482.

13. White JG, Southgate J, Thomson JN, Brenner FRS. 1986. The structure of the nervous system of the nematode *Caenorhabditis elegans*. Phil Trans Royal Soc London. Series B, Biol Scien 314(1165):1-340.

14. White J. 1988. The Anatomy. In: The Nematode *Caenorhabditis elegans* (Wood WB, ed). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 81-122.

15. Hall DH, Altun ZF. 2008. *C. elegans* atlas. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

16. Baugh LR, Hill AA, Slonim DK, Brown EL, Hunter CP. 2003. Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. Development 130(5):889-900.

17. Melstrom PC, Williams PL. 2007. Reversible AChE inhibitors in *C. elegans* vs. rats, mice. Biochem Biophys Res Commun 357(1):200-205.

18. Faber PW, Voisine C, King DC, Bates EA, Hart, AC. 2002. Glutamine/proline-rich PQE-1 proteins protect *Caenorhabditis elegans* neurons from huntingtin polyglutamine neurotoxicity. Proc Natl Acad Sci U S A 9(26): 17131-17136.

19. Guven K, Power RS, Avramides S, Allender R, de Pomerai DI. 1999. The toxicity of dithiocarbamate fungicides to soil nematodes, assessed using a stress-inducible transgenic strain of *Caenorhabditis elegans*. J Biochem Mol Toxicol 13(6):324-333.

20. Treinin M, Chalfie M. 1995. A mutated acetylcholine receptor subunit causes neuronal degeneration in *C. elegans*. Neuron 14(4):871-877.

21. Williams GT, Morimoto RI. 1990. Maximal stress-induced transcription from the human HSP70 promoter requires interactions with the basal promoter elements independent of rotational alignment. Mol Cell Biol 10(6):3125-3136.

22. Nass R, Hall DH, Miller DM, 3rd, Blakely RD. 2002. Neurotoxin-induced degeneration of dopamine neurons in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 99(5):3264-3269.

23. Aballay A, Ausubel FM. 2002. *Caenorhabditis elegans* as a host for the study of host-pathogen interactions. Curr Opin Microbiol 5(1):97-101.

24. Hoss S, Henschel T, Haitzer M, Traunspurger W, Steinberg CE. 2001. Toxicity of cadmium to *Caenorhabditis elegans* (Nematoda) in whole sediment and pore water--the ambiguous role of organic matter. Environ Toxicol Chem 20(12):2794-2801.

25. Anderson GL, Boyd WA, Williams PL. 2001. Assessment of sublethal endpoints for toxicity testing with the nematode *Caenorhabditis elegans*. Environ Toxicol Chem 20(4):833-838.

26. Marroquin LD, Elyassnia D, Griffitts JS, Feitelson JS, Aroian RV. 2000. *Bacillus thuringiensis* (Bt) toxin susceptibility and isolation of resistance mutants in the nematode *Caenorhabditis elegans*. Genetics 155(4):1693-1699.

27. Dhawan R, Dusenbery DB, Williams PL. 1999. Comparison of lethality, reproduction, and behavior as toxicological endpoints in the nematode *Caenorhabditis elegans*. J Toxicol Environ Health A 58(7):451-462.

28. Freedman JH, Slice LW, Dixon D, Fire A, Rubin, CS. 1993. The novel metallothionein genes of *Caenorhabditis elegans*. Structural organization and inducible, cell-specific expression. J Biol Chem 268(4):2554-2564.

29. Giglio AM, Hunter T, Bannister JV, Bannister WH, Hunter G. 1994. The copper/zinc superoxide dismutase gene of *Caenorhabditis elegans*. Biochem Mol Biol Int 33(1):41-44.

30. Stringham EG, Jones D, Candido EP. 1992. Expression of the polyubiquitin-encoding gene (ubq-1) in transgenic *Caenorhabditis elegans*. Gene 113(2):165-173.

31. Heschl MF, Baillie DL. 1990. The HSP70 multigene family of *Caenorhabditis elegans*. Comp Biochem Physiol B 96(4):633-637.

32. Weston K, Yochem J, Greenwald I. 1989. A *Caenorhabditis elegans* cDNA that encodes a product resembling the rat glutathione S-transferase P subunit. Nucleic Acids Res 17(5):2138.

33. Wolf M, Nunes F, Henkel A, Heinick A, Paul RJ. 2008. The MAP kinase JNK-1 of *Caenorhabditis elegans*: location, activation, and influences over temperature-dependent insulin-like signaling, stress responses, and fitness. J Cell Physiol 214(3):721-729.

34. Han M, Sternberg PW. 1990. let-60, a gene that specifies cell fates during *C. elegans* vulval induction, encodes a ras protein. Cell 63(5):921-931.

35. Lee JI, Dhakal BK, Lee J, Bandyopadhyay J, Jeong SY, Eom SH, et al. 2003. The *Caenorhabditis elegans* homologue of Down syndrome critical region 1, RCN-1, inhibits multiple functions of the phosphatase calcineurin. J Mol Biol 328(1):147-156.

36. Engvall E. 1994. Cell adhesion in muscle. Braz J Med Biol Res 27(9):2213-2227.

37. Sambongi Y, Wakabayashi T, Yoshimizu T, Omote H, Oka T, Futai M. 1997. *Caenorhabditis elegans* cDNA for a Menkes/Wilson disease gene homologue and its function in a yeast CCC2 gene deletion mutant. J Biochem (Tokyo) 121(6):1169-1175.

38. Nakao A, Roijer E, Imamura T, Souchelnytskyi S, Stenman G, Heldin CH, et al. 1997. Identification of Smad2, a human Mad-related protein in the transforming growth factor beta signaling pathway. J Biol Chem, 272(5):2896-2900.

39. Anderton BH. 1999. Alzheimer's disease: clues from flies and worms. Curr Biol 9(3):R106-109.

40. Hutton M, Hardy J. 1997. The presenilins and Alzheimer's disease. Hum Mol Genet 6(10):1639-1646.

41. Barr MM, Sternberg PW. 1999. A polycystic kidney-disease gene homologue required for male mating behaviour in *C. elegans*. Nature 401(6751):386-389.

42. Aboobaker AA, Blaxter ML. 2000. Medical significance of *Caenorhabditis elegans*. Ann Med 32(1):23-30.

43. Chen LC, Manjeshwar S, Lu Y, Moore D, Ljung BM, Kuo WL, et al. 1998. The human homologue for the *Caenorhabditis elegans* cul-4 gene is amplified and overexpressed in primary breast cancers. Cancer Res 58(16):3677-3683.

44. Pause A, Lee S, Worrell RA, Chen DY, Burgess WH, Linehan WM, et al. 1997. The von Hippel-Lindau tumor-suppressor gene product forms a stable complex with human CUL-2, a member of the Cdc53 family of proteins. Proc Natl Acad Sci U S A 94(6):2156-2161.

45. Bates GP, Hockly E. 2003. Experimental therapeutics in Huntington's disease: are models useful for therapeutic trials? Curr Opin Neurol 16(4):465-470.

46. Rubinsztein DC. 2002. Lessons from animal models of Huntington's disease. Trends Genet 18(4):202-209.

47. Link CD, Taft A, Kapulkin V, Duke K, Kim S, Fei Q, et al. 2003. Gene expression analysis in a transgenic *Caenorhabditis elegans* Alzheimer's disease model. Neurobiol Aging 24(3):397-413.

48. Kurz CL, Ewbank JJ. 2003. *Caenorhabditis elegans*: an emerging genetic model for the study of innate immunity. Nat Rev Genet 4(5):380-390.

49. Giglio MP, Hunter T, Bannister JV, Bannister WH, Hunter GJ. 1994. The manganese superoxide dismutase gene of *Caenorhabditis elegans*. Biochem Mol Biol Int 33(1):37-40.

50. Zhen M, Heinlein R, Jones D, Jentsch S, Candido EP. 1993. The ubc-2 gene of *Caenorhabditis elegans* encodes a ubiquitin- conjugating enzyme involved in selective protein degradation. Mol Cell Biol 13(3):1371-1377.

51. Candido EP, Jones D, Dixon DK, Graham RW, Russnak RH, Kay RJ. 1989. Structure, organization, and expression of the 16-kDa heat shock gene family of *Caenorhabditis elegans*. Genome 31(2):690-697.

52. Broeks A, Gerrard B, Allikmets R, Dean M, Plasterk RH. 1996. Homologues of the human multidrug resistance genes MRP and MDR contribute to heavy metal resistance in the soil nematode *Caenorhabditis elegans*. EMBO J 15(22):6132-6143.

53. Taub J, Lau JF, Ma C, Hahn JH, Hoque R, Rothblatt J, et al. 1999. A cytosolic catalase is needed to extend adult lifespan in *C. elegans* daf-C and clk-1 mutants. Nature 399(6732):162-166.

54. Ahringer J. 1997. Turn to the worm! Curr Opin Genet Dev 7(3):410-415.

55. Collins FS, Gray GM, Bucher JR. 2008. Toxicology. Transforming environmental health protection. Science 319(5865):906-907.

56. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95(1):5-12.

57. Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect 116(3):284-291.

58. Pulak R. 2006. Techniques for analysis, sorting, and dispensing of *C. elegans* on the COPAS flow-sorting system. Methods Mol Biol 351:275-286.

59. Avery L, Shtonda BB. 2003. Food transport in the *C. elegans* pharynx. J Exp Biol 206(Pt 14):2441-2457.

60. Boyd WA, Williams PL. 2003. Comparison of the sensitivity of three nematode species to copper and their utility in aquatic and soil toxicity tests. Environ Toxicol Chem 22(11):2768-2774.

61. Smith MV, Boyd WA, Kissling GE, Rice JR, Snyder DW, Portier CJ, et al. 2009. A discrete time model for the analysis of medium-throughput *C. elegans* growth data. PLoS One 4(9):e7018.

62. Boyd WA, Smith MV, Kissling GE, Rice JR, Snyder DW, Portier, et al. 2009. Application of a mathematical model to describe the effects of chlorpyrifos on *Caenorhabditis elegans* development. PLoS One 4(9):e7024.

63. Boyd WA, McBride SJ, Rice JR, Snyder DW, Freedman JH. 2010. A high-throughput method for assessing chemical toxicity using a *Caenorhabditis elegans* reproduction assay. Toxicol Appl Pharmacol 245(2):153-159.

64. Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118(4):485-492.

65. Zhu H, Ye L, Richard A, Golbraikh A, Wright FA, Rusyn I, et al. 2009. A novel two-step hierarchical quantitative structure-activity relationship modeling work flow for predicting acute toxicity of chemicals in rodents. Environ Health Perspect 117(8):1257-1264.

66. National Research Council (U.S.). Committee on Developmental Toxicology. 2000. Scientific frontiers in developmental toxicology and risk assessment. National Academy Press, Washington, DC.

IV.3 Tox21 Activities:

Probing Mechanisms of Inter-individual Susceptibility to Toxicants with Population-based Experimental Approaches

This page left intentionally blank

IV.3 Probing Mechanisms of Inter-individual Susceptibility to Toxicants with Population-based Experimental Approaches

Ivan Rusyn, M.D., Ph.D.
Director Laboratory of Environmental Genomics, Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill

Associated NTP Biomolecular Screening Branch Scientists
John E. (Jef) French, Ph.D.
Kristine Witt, M.S.

Collaborating Scientists
Christopher Austin, M.D. (NHGRI/NCGC)
Ruili Huang, Ph.D. (NHGRI/NCGC)
David Threadgill, Ph.D. (North Carolina State University)
Fred Wright, Ph.D. (University of North Carolina at Chapel Hill)
Menghang Xia, Ph.D. (NHGRI/NCGC)

IV.3.1 Background and Rationale

Genetic polymorphisms have a profound effect on differences between individuals who may have developed disease after exposure to toxicants, yet these factors are not being fully considered in safety evaluation or risk assessment. Indeed, the need to account for differences among humans in susceptibility to adverse effects of chemicals, other than from possible early-life susceptibility, is becoming ever more evident to both the scientific community and the regulatory agencies. In addition, as the risk assessment process is shifting towards greater reliance on the *in vitro* data, the quantitative assessment of the extent of inter-individual variability in responses to chemicals and understanding of its genetic causes are needed, so that the evaluation of uncertainty can be based on solid science, not defaults.

Elucidation of the genetic determinants for inter-individual differences in toxicity may be conducted using various *in vivo* (e.g., panels of inbred mouse strains) or *in vitro* (e.g., mouse or human cells that have been densely genotyped) approaches. Furthermore, the toxicity phenotypes collected in the "population-based" models may be combined with genetics and gene expression datasets, which may offer valuable insights into the molecular mechanisms for genetically-determined variability in responses, and provide necessary science-based underpinnings and tools for improved toxicity testing and risk assessment.

IV.3.2 Key Issue, Hypothesis Tested, or Problem Addressed

Our work aims to apply novel toxicogenetic tools to address the issue of inter-individual variability in susceptibility to environmental agents. The need to "account for differences among humans in cancer susceptibility other than from possible early-life susceptibility" (National Research Council 2008) is becoming ever more evident to both the scientific community and the regulatory agencies. This work also addresses the goal of the NIEHS Office of the Director to "assist in achieving the NTP Vision

&Roadmap for future research, particularly contributing to the development of new animal models of genetic susceptibility," and is well in line with the NIEHS strategic goals (NIEHS 2006).

IV.3.3 Approach/Results/Progress

We utilize three primary research tools which combine *in vivo* and *in vitro* toxicogenetic approaches: (i) a genetically-defined panel of human lymphoblastoid cell lines, (ii) toxicity studies using large panels of genetically-diverse inbred mouse strains, and (iii) isolated cultured primary hepatocytes from genetically-diverse inbred mouse strains.

IV.3.3.1 Human Lymphoblastoid Cell Lines

Immortalized human lymphoblastoid cell lines have been used to demonstrate that it is possible to use an *in vitro* model system to identify genetic factors that affect responses to xenobiotics. To extend the application of such studies to investigative toxicology by assessing inter-individual and population-wide variability and heritability of chemical-induced toxicity phenotypes, we are using cell lines from the Centre d'Etude du Polymorphisme Humain (CEPH) trios assembled by the HapMap Consortium (Meucci et al. 2005). Our goal is to aid in the development of predictive human *in vitro* genetics-anchored models of chemical-induced toxicity.

In our first experiment, cell lines from the CEPH trios were exposed to three concentrations of 14 environmental chemicals. We assessed ATP production and Caspase-3/7 activity 24 hours after treatment. Replicate analyses were used to evaluate experimental variability and classify responses. We showed that variability of response across the cell lines exists for some, but not all chemicals, with perfluorooctanoic acid and phenobarbital eliciting the greatest degree of inter-individual variability. While the data for the chemicals used here does not show evidence for broad-sense heritability of toxicity response phenotypes, candidate genetic factors contributing to the variability in response to perfluorooctanoic acid were investigated using genome-wide association analysis.

In our second experiment, we partnered with NTP and NCGC to test the population-based quantitative high throughput (qHTS) screening paradigm, in which hundreds of compounds may be profiled rapidly in dozens of cell lines and multiple biological targets, is one of the major data streams in computational toxicology (Xia et al. 2008). While existing qHTS approaches have been applied to a number of human and rodent cell lines, thus allowing for species and tissue comparisons, the inter-individual genetic variability has not been considered at this scale yet. Lymphoblast cell lines from 27 CEPH trios were exposed to 240 environmental chemicals in 12 concentrations (from 0.26 nM to 46.0 μM). We assessed caspase-3/7 activity, a marker of apoptosis, and intracellular ATP production, a measure of cell viability, and 16 and 40 hours, respectively, after treatment, with 2-3 replicates per concentration tested. qHTS screening in the genetically-defined population of human cells produced robust and reproducible results, which allow for cross-compound, -assay and -individual comparisons. Some compounds were cytotoxic to all cell types at similar concentrations, whereas others exhibited inter-individual differences in cytotoxicity.

IV.3.3.2 Toxicity Studies

To enhance the utility of animal models to uncover mechanisms of toxicity and detect response biomarkers in genetically diverse populations, we use genetically-defined panels of mouse strains (Rusyn et al. 2010). By taking into the account strain-specific chemical metabolism, toxicity phenotypes, and gene expression patterns, it is possible to establish genetic polymorphism-dependent and -independent pathways perturbed by the toxicants. This approach has been applied in three case studies where liver

effects of acetaminophen (Harrill et al. 2009a, Harrill et al. 2009b), trichloroethylene, butadiene and ethanol were assessed. Since genetic regulation of gene expression is a key contributor to population diversity (Gatti et al. 2007, Gatti et al. 2009), these studies provide better understanding of the mechanisms of toxicity that may define susceptibility or resistance.

IV.3.3.3 Primary Hepatocytes

In our studies with cultures of hepatocytes obtained from a large panel of inbred mouse strains we aimed to generated data indicative of inter-individual differences in *in vivo* responses to hepatotoxicants (Martinez et al. 2010). In order to test this and establish whether *in vitro* studies using cultured hepatocytes from genetically distinct mouse strains are feasible, we aimed to determine whether viable cells may be isolated from different mouse inbred strains, evaluate the reproducibility of cell yield, viability and functionality over subsequent isolations, and assess the utility of the model for toxicity screening. Hepatocytes were isolated from 15 strains of mice and cultured for up to 7 days in traditional 2-dimesional culture. Cells from 3 strains were treated with acetaminophen, WY-14,643, or rifampin and concentration-response effects on viability and function were established. Our data suggest that high yield and viability can be achieved across a panel of strains. Cell function and expression of key liver specific genes of hepatocytes isolated from different strains and cultured under standardized conditions is comparable. Strain-specific responses to toxicant exposure have been observed in cultured hepatocytes and these experiments open new opportunities for further developments of *in vitro* models of hepatotoxicity in a genetically diverse population.

IV.3.4 Significance

This work represents interdisciplinary (toxicology, genetics, biostatistics, pharmacokinetic modeling) research aimed at elucidating the genetic basis of dose-response and susceptibility; uses state-of-the-art technologies; develops new models and analysis tools for systems biology approaches; identifies potential biomarkers linked to genetic differences in toxicant metabolism and/or response; and generates knowledge directly applicable to quantitative risk assessment.

A shift in toxicity testing from traditional *in vivo* to higher throughput *in vitro* methods creates promise to prioritize compounds, uncover mechanisms of toxicity and provide rich data for predictive modeling of adverse health effects. The approach of screening chemicals for toxicity in a genetically-defined, yet diverse *in vitro* rodent and human cell-based system is potentially useful for identification of both chemicals and individuals that may be at highest risk, the extent of within-species variability in the population, and genetic loci of interest that potentially contribute to chemical susceptibility. The generation of high-quality cytotoxicity data on large libraries of compounds using qHTS demonstrates the potential of this methodology to profile a much broader array of assays and compounds while exploring the genetic determinants of inter-individual variability and potentially predicting *in vivo* biological response.

IV.3.5 References

Gatti D, Maki A, Chesler EJ, Kirova, R, Kosyk O, Lu L, et al. 2007. Genome-level analysis of genetic regulation of liver gene expression networks. Hepatology 46:548-557.

Gatti DM, Harrill AH, Wright FA, Threadgill DW, Rusyn I. 2009. Replication and narrowing of gene expression quantitative trait loci using inbred mice. Mamm Genome 20:437-446.

Harrill AH, Ross PK, Gatti DM, Threadgill DW, Rusyn I. 2009a. Population-based discovery of toxicogenomics biomarkers for hepatotoxicity using a laboratory strain diversity panel. Toxicol Sci 110:235-243.

Harrill AH, Watkins PB, Su S, Ross PK, Harbourt DE, Stylianou IM, et al. 2009b. Mouse population-guided resequencing reveals that variants in CD44 contribute to acetaminophen-induced liver injury in humans. Genome Res 19:1507-1515.

Martinez SM, Bradford BU, Soldatow VY, Kosyk O, Sandot A, Witek R, et al. 2010. In press. Evaluation of an in vitro toxicogenetic mouse model for hepatotoxicity. Toxicol Appl Pharmacol.

Meucci MA, Marsh S, Watters JW, McLeod HL. 2005. CEPH individuals are representative of the European American population: implications for pharmacogenetics. Pharmacogenomics 6:59-63.

National Research Council. 2008. Science and Decisions: Advancing Risk Assessment. The National Academies Press, Washington, D.C.

NIEHS (National Institute of Environmental Health Sciences). 2006. New frontiers in environmental sciences and human health: The 2006-2011 NIEHS Strategic Plan. Available: http://www.niehs.nih.gov/about/od/strategicplan/ [accessed 16 October 2010].

Rusyn I, Gatti DM, Wilshire T, Kleeberger SR, Threadgill DW. 2010. Toxicogenetics: population-based testing of drug and chemical safety in mouse models. Pharmacogenomics 11:1127-1136.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect 116:284-291.

IV.4 Tox21 Activities:

Mining NTP Tissue Archives for Gene Signatures

This page left intentionally blank

IV.4 Mining NTP Tissue Archives for Gene Signatures

B. Alex Merrick, Ph.D.
Biomolecular Screening Branch

Biomolecular Screening Branch Scientists
Scott S. Auerbach, Ph.D., DABT

Associated NTP Scientists
Julie F. Foley (CMPB)
G. Jean Harry, Ph.D. (NLB)
Mark J. Hoenerhoff, D.V.M., Ph.D. (CMPB)
Patricia S. Stockton (CMPB)
David E. Malarkey, D.V.M (CMPB)
Robert C. Sills, D.V.M., Ph.D. (CMPB)
Erik J. Tokar, Ph.D. (NLB)
Michael P. Waalkes, Ph.D. (NLB)

Associated NIH/NIEHS Scientists
NIEHS NextGen Sequencing Committee
NIH Intramural Sequencing Center (Gaithersburg, MD)

Collaborating Scientists
Yuan Gao, Ph.D. (Johns Hopkins University, Baltimore, MD)
Arun K. Pandiri, D.V.M., Ph.D. (Environmental Pathology Laboratories, Inc., Research Triangle Park, NC)
Ruchir R. Shah, Ph.D. (SRA International)
David H. Murdock Research Institute (Kannapolis, NC)

IV.4.1 Introduction

The primary purpose of this project is to evaluate the extent to which gene expression signatures can be reliably derived from the molecular analysis of tissue samples collected from the laboratory animals used in NTP's toxicological studies and stored as formalin fixed, paraffin embedded (FFPE) tissues in the NTP archives. Signature expression profiles can be described as a critical set of up- and down-regulated transcripts that can distinguish between health, toxicity, and disease (Merrick and Bruno 2004). Development of such signatures would further our understanding of pathological changes and mechanisms of agent-induced toxicity, as well as aid in the identification of "toxicity" pathways. These signatures would therefore contribute to the identification of useful targets for *in vitro* assays, to an evaluation of the correlation between *in vitro* test results and *in vivo* toxicological outcomes, and to the development of predictive models of toxicity.

For many of the toxicological studies conducted by the NTP, frozen tissues are not always available for extraction of full-length RNA. Assuming that the challenges of using less than optimal RNA extracted from FFPE tissue can be overcome, performing molecular analyses on archival paraffin block tissues would greatly expand our ability to link changes in gene expression with disease outcomes while leveraging the considerable expense already invested by the U.S. government in the more than 25 years of high quality toxicological studies conducted by the NTP.

The FFPE tissues stored in the NTP archives represent a significant and underutilized resource. In recognition of this, we have initiated studies to evaluate different technologies for their ability to be applied to FFPE tissues for generating gene signatures of agent-induced toxicity.

The NTP archives were established in 1984 and contain stained histopathology slides, paraffin tissue blocks, formalin fixed sealed tissues and organs, and fresh frozen tissues stored in liquid nitrogen. Samples from over 2,000 NTP studies are housed in these archives; these include toxicity, carcinogenicity, immunotoxicity, reproductive, and developmental studies. Study results and images are stored in print, microfiche, and digitized formats. The frozen tissue banks contain samples from 169 toxicity and carcinogenicity studies. Frozen samples include normal tissue, non-neoplastic lesions, tumor specimens, DNA, RNA, blood serum, bronchial alveolar lavage fluid, urine, and sperm suspensions from treated and control rats and mice. Overall, the NTP archives contain:

- >2,000 NTP studies
- >7.5 million histological glass slides
- >4.6 million paraffin-embedded tissue blocks
- >230 thousand sealed bags of formalin-preserved tissue
- >54 thousand frozen specimens
- Histopathology images that include >50 thousand 2x2 kodachrome slides and >20,000 digital images
- Study data that include >3.5 million pages, >10 million pages of microfiche data, >1.5 million pages of digital or electronic records

Holdings of the NTP archives continue to expand. In 2009, for example, 80,203 histology glass slides, 72,200 paraffin blocks, 7,590 formalin tissues, and 1,338 frozen specimens were added to the archives. These data indicate that the NTP archives are an actively growing collection for a wide variety of tissues sampled from studies involving many different chemical and physical agent exposure conditions. The NTP archives are used by both intra- and extra-mural scientists; within the same year (2009), there were 609 NTP, 50 DIR, and 21 outside individual requests for archival samples. The ability to mine archival samples for gene signatures of toxicity at both the discovery and validation levels would add an important dimension to an already important NTP resource.

IV.4.2 qPCR Analysis of FFPE Liver of Rats Exposed Subchronically to Aflatoxin B1 for Gene Signatures

IV.4.2.1 Background and Rationale

Earlier this year, NTP scientists published the results of a liver microarray study conducted to predict the hepatocarcinogenic potential of alkenylbenzene flavoring agents in male Fischer 344 rats using toxicogenomics and machine learning (Auerbach et al. 2010). Aflatoxin B_1 (AFB1), a genotoxic potent liver carcinogen, was included among the known carcinogens used to train SVM models of gene expression developed to identify chemicals with hepatocarcinogenic activity in rats. AFB1 is one of the most positive of all NTP chemicals for rat liver cancer, causing a >90% incidence of multiple HCCs in male rats in a two-year bioassay. Agilent rat whole genome oligonucleotide microarrays in a 4x44K format were used to evaluate alterations in gene expression using RNA isolated from liver. A number of genes informing SVM prediction models were found to be differentially expressed at 90 days of exposure to 1 ppm AFB1, prior to development of liver tumors, altered histopathology, or altered serum chemistries. Several genes were found to be upregulated by exposure to AFB1; these included the oncogene, Mybl2

(Sala 1999), the transporters, Abcb1b and Abcc3 (Klaassen et al. 2010), and the tissue remodeling metalloproteinase, Adam8 (Murphy 2008). The up-regulation of these genes was accompanied by down-regulation of the tumor suppressor genes, Wwox and Fhit (Iliopoulos et al. 2006).

IV.4.2.2 Key Issue, Hypothesis Tested, or Problem Addressed

A pilot study was conceived to determine if selective transcript amplification by quantitative polymerase chain reaction (qPCR) was possible using RNA from NTP archival FFPE tissue. Tissue blocks would be used that represented standard fixation protocol, embedding procedures, and storage in the NTP archives. The recently conducted NTP microarray gene expression study described above was selected for this study. The hypothesis was that qPCR of FFPE RNA from AFB1 and control liver would provide gene fold changes that were quantitatively comparable to those generated from Agilent microarrays.

IV.4.2.3 Approach

Studies were performed to determine:

1. an acceptable commercial method for extracting RNA
2. if RNA could be extracted from NTP archival FFPE samples in sufficient amounts for target validation, and
3. if RNA was of sufficient quality for transcript analysis.

In collaboration with Julie Foley, Patricia Stockton, and Robert Sills of the NTP CMPB, we determined that the PureLink Extraction kit by Invitrogen produced an acceptable amount of RNA from archival FFPE samples. Samples were digested with DNase I prior to reverse transcriptase amplification with random hexamers. Both S18 or β-actin served as representative, housekeeping genes, producing low cycle threshold (Ct) values and acceptable A260/280 ratios from 1.9 to 2.0. These preliminary studies showed initial criteria of quantity, quality, and qPCR amplification could be met using RNA isolated from FFPE sample blocks.

IV.4.2.4 Results/Progress

Previous microarray analysis of liver showed that a 90-day exposure of male F344 rats to 1 ppm AFB1 in feed altered the expression of hundreds of genes prior to liver tumor development or detectable histopathological changes (Auerbach et al. 2010). We evaluated the expression of 11 genes whose selection was based on microarray data by qPCR using RNA extracted from FFPE liver prepared from the same rats (6 control and 6 AFB1 treated rats per group) used in the microarray study. RNA extraction from four combined sections/archival liver paraffin block yielded 40 to 50 µg RNA that ranged in size from 0.1 to 4 kB measured by Bioanalyzer. qPCR showed increased fold expression in AFB1 treated rats compared to controls of 10X for the oncogene, Mybl2; 32X and 23X, respectively, for the drug transporters, Abcb1b and Abcc3; 91X for the disintegrin metalloprotease, Adam8; 190X for DNA damage-inducible, Ddit4L; 54X for the glutamate receptor subunit, Grin2c; 11X for xenobiotic-responsive, uncharacterized protein c8orf46; and 65X for the cadherin, Cdh13. For the tumor suppressor genes, Fhit expression was lower by 1.5X than control and Wwox was unchanged. Expression of the chemokine, Cinc1, was unchanged.

IV.4.2.5 Significance

These data indicate that RNA can be isolated from NTP FFPE archival tissues in sufficient quantity and quality for comparative gene expression by qPCR. Overall, expression data for select genes from qPCR and microarray platforms agreed well for directionality and fold change. By qPCR, 5 genes exhibited higher expression responses, 4 genes exhibited similar responses, and 1 gene showed a lower response compared to microarray. This study paves the way for further directed inquiries into NTP archival tissues for signature validation and potentially for signature discovery as well.

IV.4.3 Comparative Global Gene Expression Analysis of NTP Fresh Frozen and FFPE Archival Tissues using NextGen Sequencing

IV.4.3.1 Background and Rationale

Hybridization based arrays have been the central technology in toxicogenomic profiling since their introduction more than a decade ago (Nuwaysir et al. 1999). However, 'NextGen' sequencing represents a series of new technologies for massive parallel DNA sequencing that provides more information about the transcriptome with better accuracy than hybridization-based arrays, the current standard for expression analysis (Mardis et al. 2008). Whole transcript analysis by NextGen sequencing is termed RNA-Seq. After isolation and fragmentation of mRNA, a cDNA library of nucleotide sequences is created from which millions of short DNA reads are generated in single or paired-end orientation (Langmead et al. 2009). After alignment to a reference genome, expression of each portion of a gene expressing a transcript is counted from aligned reads (Simpson et al. 2009, Trapnell et al. 2009). Fold changes can be made relative to a designated control. RNA-Seq has advantages of a higher dynamic range at >9000 fold compared to hybridization array (~100 fold range); more sensitive detection of low copy transcripts; measurement of 5' and 3' splice variants; unbiased mapping of intron-exon boundaries; and determination of premature stop codons, indels, and non-coding, untranslated regions (Sampath et al. 2007, Trapnell et al. 2010). Thus, a more detailed description of the transcriptome is possible with NextGen approaches.

One common drawback to the use of gene expression microarrays and NextGen sequencing approaches is that gene expression is best determined from fresh frozen tissue but such samples, in contrast to FFPE tissue blocks, have not been routinely archived by the NTP. Protocols for RNA-Seq typically depend upon separation of mRNA, which comprises only 2-5% of total RNA, from the remainder comprised mainly of ribosomal RNA. Selection for non-ribosomal transcripts is based on either synthesizing cDNA using oligo-dT primers followed by fragmentation of cDNA or by initially selecting for polyA-tailed RNA followed by RNA fragmentation and cDNA synthesis using random hexamer priming. These approaches depend upon starting with high quality total RNA which is typically not the case when isolating RNA from FFPE tissue samples.

Overcoming the limitations of degraded RNA for use in gene expression profiling studies with clinical FFPE samples was the focus of a new method developed by researchers at Stanford University (Beck et al. 2010). They recently reported development and validation of a technique called 3'Seq, to apply the power of NextGen sequencing to degraded RNA extracted from paraffin blocks. The distinguishing feature of 3'Seq from standard RNA-Seq protocols is that while RNA-Seq generates a non-directional sequencing library of RNA fragments that span the entire transcript length, the 3'Seq protocol generates a directional sequencing library comprised primarily of 200 bp cDNA fragments with a poly-A tail. Sequencing proceeds directionally toward the poly-A tail. This design allows more accurate transcriptional profiling from degraded RNA since it ensures that at least one read per transcript

molecule will be produced regardless of transcript length. While 3'Seq does not provide information on splice variants, it can deliver a more sensitive and precise profile (lower false discovery rate, more sensitive detection of low copy transcripts) than hybridization based arrays for which degraded RNA creates hybridization artifacts and background noise.

IV.4.3.2 Key Issue, Hypothesis Tested, or Problem Addressed

We are evaluating the extent to which RNA-Seq better describes, when compared to hybridization-based microarrays, critical genes and pathways involved in chemically-induced rat liver toxicity and carcinogenesis. In addition, we are evaluating the application of 3'Seq to RNA isolated from FFPE archival rat liver tissues and using RNA isolated from the corresponding fresh frozen tissue as the standard for comparison. Importantly, since the same RNA samples from fresh frozen liver tissue are being used for RNA-Seq and 3'Seq, we will be able to directly compare the performance of these two NextGen technologies in terms of gene expression changes induced in liver by a liver carcinogen.

IV.4.3.3 Approach

High quality RNA was extracted from fresh frozen liver samples in the NTP archives prepared from concurrent control and 90-day AFB1 (1 ppm) feed treated male F344 rats (Auerbach et al. 2010). The primary goal of this part of the project is to more precisely define gene expression changes produced by AFB1 related to carcinogenesis using RNA-Seq. We hypothesize that AFB1 produces gene expression differences in splice variants, low copy transcripts, and non-coding RNA species important in understanding the development of HCCs in the rat model. In addition to widespread changes in transcript expression during the development of cancer, alternative splicing represents an important molecular mechanism of gene regulation in physiological processes such as developmental programming. Tumors express a different collection of alternative splice-forms than normal tissues, suggesting they may play a pathogenic role in cancer (Körner and Miller 2009). For example, alternative splicing of ion transporters and inhibitors of apoptosis may have a role in HCC development (Notarbartolo et al. 2004, Jonker et al. 2004). The importance of splice-forms in carcinogenesis is suggested by formation and overexpression of Mad1beta, a splice variant of the mitotic checkpoint control protein MAD1 found in patients with HCCs. This splice variant causes mitotic checkpoint impairment, chromosome bridge formation, and aberrant chromosome numbers via binding to MAD2 (Sze et al. 2008). Similarly important splicing events may play a role in chemically-induced HCC in the rat. Given the importance of AFB1 as a worldwide environmental factor in human liver cancer, more complete description of gene expression differences of splice variant, low copy transcripts, and non-coding RNA species in experimental animal models of liver cancer could contribute to better understanding of the process leading to HCC. A second goal is to establish a high resolution map of the F344 rat liver transcriptome. This objective would include defining and measuring levels of splice variants, refinement of intron-exon boundaries, and the frequency and amounts of splice variants that might be expected. RNA-Seq analysis of mouse liver, brain, and skeletal muscle shows that about 3% of all reads were splice-spanning reads, in which more than one alternate splice form was observed for 3,462 genes in the three tissues analyzed (Mortazavi et al. 2008). This same study also suggests that ample opportunity for improved transcript annotation and novel transcript discovery would likely occur in the rat, since RNA-Seq of the three mouse tissue transcriptomes found 596 novel candidate transcripts (if protein-coding was not determined). Further refinement of intron-exon boundaries and detection of rare transcripts could also be obtained, features not usually observed by hybridization arrays. A high resolution map of the liver transcriptome from the male F344 rat is not yet available but defining it would be important to the NTP program since this strain is frequently used in many experimental and preclinical studies.

The second but related part of this project is being performed to evaluate 3'Seq on NTP archival samples in collaboration with Dr. Yuan Gao at the Johns Hopkins University in Baltimore, MD. Total RNA has been extracted from both fresh frozen and FFPE liver samples from concurrent control and AFB1 treated male F344 rats (same study as above). 3'Seq analysis of fresh frozen liver will serve as the primary data comparison for 3'Seq gene expression acquired from FFPE liver samples from the same animals. Both sets of 3'Seq data will be compared to fresh frozen RNA-Seq data for similarities and differences in global gene expression.

IV.4.3.4 Results/Progress

A project proposal was submitted and approved by the NIEHS Next Generation Sequencing Committee (July 2010) for RNA-Seq analysis of RNA extracted from fresh frozen concurrent control and AFB1 livers. NextGen sequencing is being performed on an Illumina IIx instrument by the NIH Intramural Sequencing Center (NISC). The first pair of samples (one control, one AFB1-treated rat) has been sequenced and is currently undergoing bioinformatic analysis by Dr. Ruchir Shah, a NextGen informatics specialist at NIEHS. Other pairs of control and AFB1 samples will be sequenced in the coming months.

For the 3'Seq study, procedures for polyA capture, adaptor ligation, and library construction are being performed in collaboration with Dr. Yuan Gao, an expert in the NextGen field. Samples were shipped to Dr. Gao in October and are being analyzed on an Illumina HiSeq2000.

IV.4.3.5 Significance

These studies will determine if NextGen sequencing (RNA-Seq and 3'Seq) provides a more detailed and useful description of gene expression changes in archived fresh frozen liver samples from rats exposed to AFB1 versus control rats, when compared to prior DNA microarray studies using the same samples. Also, the 3'Seq technology will be evaluated for its ability to detect gene expression changes in FFPE liver compared to corresponding fresh frozen tissue from the same animals. In addition, the transcriptome in the male F344 rat liver will be more completely described by expression of exomic regions and for detection of non-coding transcripts.

IV.4.4 Quantitative Nuclease Protection Assay (qNPA) for Gene Signature Validation in a Medium Throughput, Multiplexed Platform

IV.4.4.1 Background and Rationale

DNA microarray analysis of selected samples from NTP rodent studies can produce transcript profiles that are important for describing the underlying pathology in terms of gene expression changes. Bioinformatic analysis can often condense these large datasets into small clusters of critical gene changes – gene signatures – that might be predictive of chemical-induced toxicity and cancer. Validation for small gene sets using full- or large-scale transcriptome probes sets by microarray is impractical because of low throughput and expense. Thus, the validation of gene expression signatures and their application to various tissues and species would be greatly assisted by a flexible, medium throughput, multiplexed platform. In addition, validation of gene signatures to test their prediction potential should not be limited to the availability of fresh frozen tissue but ideally should extend to FFPE tissue samples. Due to the degraded nature of RNA in FFPE tissues, special processing of paraffin samples must be combined with highly sensitive techniques for measuring multiple genes at the same time for each sample in a multiplexed mode.

The need for a medium throughput, multiplexed platform for gene expression signature validation can be met by the qNPA platform in 96-well or 384-well formats. This platform allows for:

- a customizable selection of gene probes specific for the species of interest. Probes can be based upon public database information using NCBI Accession numbers
- sample preparation that accommodates analysis of RNA from FFPE samples as well as from fresh frozen tissue
- gene expression analysis that is amenable to using high quality RNA from fresh frozen tissue or poorer quality RNA from FFPE tissue
- the analysis of multiple genes; in 96-well and 384-well format, each sample well can contain 47 or 9 specific gene expression probes, respectively, for simultaneous analysis
- normalization of signals to housekeeping genes that are not affected by treatment and can be measured in each well
- high sensitivity for differential gene expression resolution at gene fold changes of 20% or less

IV.4.4.2 Key Issue, Hypothesis Tested, or Problem Addressed

The qNPA platform makes possible the rapid analysis of gene signatures, depending on the number of genes interrogated, from up to 96 or 384 samples at the same time. The NTP will determine if the 96-well qNPA platform is widely applicable for use in validating gene signatures in fresh frozen and FFPE tissues in mice in three different NTP-related contexts. NTP scientists have selected genes that will comprise the qNPA probes to be evaluated based on previous microarray expression data. These studies will be conducted:

- to determine if HCC and lung carcinoma gene signatures derived from prior DNA microarray studies of chemically-induced tumors in B6C3F1 mice are conserved in spontaneous liver and lung tumors from the same mouse strain (Dr. Hoenerhoff).
- to determine if gene expression signatures in the trimethyltin (TMT)-induced hippocampus damage model, as derived from focused DNA expression arrays on hippocampus, can be observed as accentuated or diminished in specific hippocampus regions such as the dentate gyrus and the CA3 (Dr. Harry).
- to validate a neuronal signature profile based on multiple gene expression and molecular studies. The selected probes represent a neuronal gene signature of the central nervous system and peripheral nerve tissues that would query neuronal resident cell type, neuronal receptors and growth factors, synaptic structure, neural development, and neuritogenesis for use in NTP studies with known or suspect chemicals causing neuronal deficits and neurotoxicity (Dr. Harry).

In summary, the qNPA platform appears to be able to accommodate investigator-initiated selection of genes to construct custom signatures using RNA extracted from either fresh frozen or FFPE tissue and therefore potentially represents a very useful capability for rapidly analyzing tissue samples in multiplex format from ongoing NTP studies or from its archive repository.

IV.4.4.3 Results/Progress

A contract for qNPA analysis was awarded to High Throughput Genomics, Inc. in late September 2010. Samples are in the process of being prepared to ship to this company for analysis in 96-well format.

IV.4.4.4 Significance

This pilot study will evaluate the suitability of the qNPA system as a medium throughput platform for evaluation of customizable gene signatures in NTP archived fresh frozen and FFPE tissue samples. The ability to perform signature validation on either archived fresh frozen or FFPE tissues would be a great asset to the NTP. In the current qNPA platform, gene signatures of up to 47 or 9 genes can be evaluated in each well of a 96-well or 384-well format, respectively.

IV.4.5 NextGen Sequencing for Epigenetic Effects of Arsenic and Cadmium in Transformed Cells

IV.4.5.1 Background and Rationale

To increase our understanding of the capabilities of NextGen sequencing and its potential applications to toxicological research, we initiated a study to evaluate its utility for detecting epigenetic changes potentially related to carcinogenicity. Epigenetic changes that involve alterations in DNA methylation during cancer development and progression are prime contributors to differential gene expression and chemically-induced malignant transformation. Defects in DNA methylation of particular DNA sequences including hypomethylation and hypermethylation have been shown to be highly associated with various malignancies in humans and rodents. Aberrant methylation of promoter regions contributes to carcinogenesis resulting in overexpression of ERα and TFF3 (Trefoil factor 3) in HCC or inactivation of expression of tumor suppressor genes like p16/cdkn2a and RASSF1A (RAS association family 1 gene) in a variety of malignancies. To conserve resources, the initial focus of this study is on evaluating epigenetic differences in cell lines differing in tumorigenic potential.

Dr. Michael Waalkes' laboratory has been studying transformed, human cell lines with syngeneic backgrounds, including the CAsE-PE (chronic-arsenic–exposed human prostate epithelial) and CTPE (cadmium-transformed prostate epithelial) cell lines, which have been malignantly transformed with arsenic (As) and cadmium (Cd), respectively (Achanzar et al. 2001, Achanzar et al. 2002). These cells give rise to aggressive cancers when injected into mice. Transformed cell lines were each derived from the 'normal' non-tumorigenic, parental line (RWPE-1) that were human papillomavirus (HPV) 18-immortalized from normal human prostate epithelial cells. A comparison of differential methylated DNA patterns at the whole genome level in these transformed human cells could provide mechanistic insight into the epigenetic changes produced in cells by environmental agents. Further, similarities and differences in the mechanisms of carcinogenesis by As or Cd could be revealed by mapping the sites of DNA methylation by bisulfite sequencing after methyl-DNA enrichment and associating those methylation sites with gene expression by RNA-Seq.

IV.4.5.2 Key Issue, Hypothesis Tested, or Problem Addressed

We hypothesize that long-term exposure of cells to As and Cd produced site-specific changes in DNA methylation that alters gene expression to contribute to malignant transformation compared to the syngeneic, non-tumorigenic RWPE-1 parental cells. The goal is to provide data on exact sites of methylation on DNA fragments from NextGen sequencing of enriched methylated DNA fragments undergoing bisulfite treatment in control and As and Cd transformed cell lines. This study will increase

our understanding of the capabilities of NextGen sequencing and its potential applications to toxicological research.

IV.4.5.3 Approach

Strategies and tools for methylated DNA enrichment by affinity binding are commercially accessible and streamlined. NextGen bisulfite DNA sequencing will be partnered with affinity enrichment procedures for methylated DNA. By combining enrichment, bisulfite reduction, and NextGen sequencing, we would accomplish a discovery-based genomic analysis of methylated DNA without the necessity for prior sequence knowledge and targeting of discrete methylated genomic regions.

The majority of DNA methylation in mammals occurs in 5'-CpG' dinucleotides in short stretches of DNA called CpG islands. One relatively new approach for enrichment of methylated DNA from fragmented whole genomic DNA exploits specific binding to the methyl-CpG binding domain of human MBD2 (Methyl-CpG-binding domain protein 2), coupled to paramagnetic beads via a biotin linker. The methylated DNA fragments would then be eluted as a single enriched population by high salt elution. Eluted methylated DNA fragments can be bisulfite reduced prior to sequencing to determine exact sites of methylation. Gene expression accomplished by RNA-Seq would be performed on the same samples for association of the methylated regions with gene expression. Validated of methylation sites and linkage of methylation to gene specific expression can be validated in subsequent experiments in the Waalkes' laboratory through a variety of established methods.

The conceived workflow would be DNA isolation, fragmentation, methylated DNA enrichment by MBD2-beads, and bisulfite treatment followed library construction and DNA-Seq by an Illumina HiSeq2000 instrument. RNA-Seq would be performed on a parallel set of samples for gene expression.

IV.4.5.4 Results/Progress

The parental control RWPE-1 cells and As or Cd transformed cell lines (CAsE-PE and CTPE cells, respectively) have been harvested by Drs. Tokar and Waalkes. A sample of each cell line is designated for RNA-Seq and another cell set will be used for bisulfite DNA sequencing after methyl DNA enrichment. These analyses will be carried out by the David H. Murdock Research Institute. Bioinformatic analysis will be conducted by Dr. Ruchir Shah, with assistance of other NIEHS informaticians familiar with NextGen sequencing data and alignment and assembly algorithms. Bioinformatic analysis of mRNA and bisulfite reduced DNA alignments should be facilitated by the annotation of the human genome.

IV.4.5.5 Significance

First, insights into epigenetic mechanisms that are associated with As and Cd-induced transformation should be derived from differences in amount and specificity of methylation in these cells versus the parental control. Data generated from this study will have the benefit of a syngeneic background since the CAsE-PE and CTPE were both derived from the human prostatic cell RWPE-1 cell line. Dr. Waalkes' laboratory plans on appropriate follow-up studies to exploit findings from NextGen data. Second, this pilot study could serve as a model for developing genome wide epigenetic queries for NTP studies when fresh frozen tissue is available. Additional data also suggests that these same enrichment procedures for methylated DNA will be applicable to archival FFPE tissue specimens. In addition to FFPE tissues stored in the NTP archives, Dr. Waalkes has a considerable number of human and rodent FFPE samples that could also be queried by site-specific methylation for comparison with gene expression to lend further insight into arsenic-induced human cancers.

IV.4.6 Proteomics Platform for Evaluation of Fresh Frozen, Formalin Fixed, and FFPE Tissues

IV.4.6.1 Background and Rationale

A pilot study was conducted to evaluate a sensitive antibody-based proteomics platform manufactured by Zeptosens, Inc. (Basel, Switzerland). This platform offers the ability to examine protein signaling pathways implicated in toxicological responses. The purpose of the study was to examine the ability of the platform to detect both total and activated (phosphorylated) proteins in fresh frozen, formalin fixed, and FFPE tissue. Pathway activation responses were represented by changes in phospho-Akt (2 sites), phospho-CREB, CREB, phosphor-ERK1/2 and ERK. The positive control model selected was the TMT-induced hippocampal damage model that has been extensively studied by Dr. Harry's group at NIEHS. For this model, Dr. Harry has previously generated data for temporal and spatial induction of various mRNAs, proteins, and protein activation, including the specific proteins chosen for this platform for the hippocampus. Thus, by using this targeted proteomics system, we are able to directly compare the generated data from the platform to her existing data for validation and to also potentially obtain new information on protein responses in subregions (CA1 and dentate gyrus). We also examined liver responses to acute TMT exposure as a non-target tissue organ.

IV.4.6.2 Key Issue, Hypothesis Tested, or Problem Addressed

The hypothesis was that the Zeptosens antibody array proteomics platform would detect activation of the Akt, CREB, and Erk pathways by phosphorylation in specific regions of the mouse brain after acute neurotoxicant treatment. Also, the data obtained in this study would permit an evaluation of the ability of this platform to be used not only on fresh-frozen tissue but also on formalin fixed and/or FFPE tissues.

IV.4.6.3 Approach

A custom protein array analysis (Phospho-AktSer473; Phospho-AktThr308, CREB; Phosphor-CREBSer133; ERK1/2; and Phosphor-ERK1/2Thr202, Tyr204) was conducted on NTP samples (fresh frozen, formalin fixed, and FFPE) to evaluate signaling pathways activated as a toxicity response in a neurological mouse model of toxicity based on sampling mice 48 hours after treatment with TMT. Each protein sample in the array was measured in duplicate for reproducibility. Each sample had four dilutions in the array to insure a concentration-related presence of antibody reactivity with the sample as a means of minimizing false positives. The antibody array accommodated 64 cell lysates for analysis.

IV.4.6.4 Results/Progress

Results of this analysis were submitted to NTP in August 2010. Strong signals were obtained from assays using fresh frozen tissue with Erk1/2 and phospho-Erk1/2. Phospho-CREB, phosphor-CREB, and phosphor-Akt Thr308 showed TMT neurotoxicant related changes within certain regions of the hippocampus suggesting a heterogeneity of response within specialized neurons of this region of the brain. However, for formalin fixed and FFPE sections of the hippocampus, the generated antibody signal in both hippocampus and liver indicated high background that did not allow distinguishable signal to be detected. It was concluded for this platform and set of lysis conditions that this methodology would not be useful for detecting protein changes in formalin fixed or FFPE tissues.

IV.4.6.5 Significance

Pathway activation from phosphorylation of selected hub proteins in signaling pathways can indicate underlying mechanisms of chemically-induced pathologies. Use of medium throughput antibody array screens can be useful tools in discerning pathway activation at key times in the progression of toxicity as is the case in the TMT model for hippocampal injury. Although formalin fixed and FFPE samples did not produce interpretable signal due to high background, it might be possible to adjust the lysis buffer and processing steps for suitable use of formalin fixed or FFPE tissues in the future. In addition, there are other affinity-based platforms that might be more applicable to FFPE tissues. The NTP will continue to explore protein detection platforms to detect parent protein and specific modifications that reflect pathway activation, degradation, signaling, and other critical cellular processes informative of chemically-induced toxicity and pathology.

IV.4.7 NTP Acquisition of DrugMatrix®

IV.4.7.1 Background and Rationale

In late September, 2010, the NTP acquired DrugMatrix® from Entelos, Inc. DrugMatrix® is a toxicogenomics reference database, tissue archives, and informatics system originally developed by Iconix in 2007 (**Table IV.4-1**). NTP acquired this resource in order to expand our ability (as well as that of the international scientific community) to develop predictive models for toxicological effects based on gene signatures, to provide an additional tool for linking *in vitro* data to *in vivo* gene signatures and disease outcomes, and to provide additional tissue samples for NextGen-based investigations.

Table IV.4-1 DrugMatrix®

Compounds
- 8000 with Structures
- 2000 with Base Curation
- 900 with Full Curation
- 874 with Molecular Pharmacology
- 657 in Expression Studies

Gene Expression
- GE CodeLinkRU1 10K rat array >15,000
- AffymetrixRat Whole Genome Arrays; >5,000

Pathology Assays
- Histopathology
- Clinical chemistry
- Hematology
- Body and organ weights

Pharmacology Assays
- Binding, Enzyme, ADME
- 130 assays x 870 compounds

The three primary components of DrugMatrix® are a toxicogenomics database, informatics tools to query the database, and an extensive tissue library. Each component has several distinguishing features. The toxicogenomics database includes a graphics user interface that allows for rapid scoring of genomic signatures of toxicity and a framework that allows for the storage and analysis of multiple data types. The *in vivo* chemical exposure study data was based on the testing of male Sprague Dawley rats, and involved multiple dose levels administered by oral gavage for acute, subacute, and subchronic durations. These studies include extensive data on pharmacology, clinical chemistry, hematology, histology, body and organ weights, and clinical observations. Also included are toxicological and toxicogenomic (microarray) data from *in vitro* studies using rat primary hepatocytes. Test agents include U.S. FDA approved drugs, standard biochemicals, and environmental toxicants with curation of all relevant public information. Advanced query tools allow access to all information in integrated data domains. Queries

can combine literature-derived information on genes or compounds, pharmacological activity, expression changes, or chemical structure to identify agents that modulate, induce, or repress specific genes. Affected biochemical pathways, specific mechanisms of action or therapeutic class can be queried. Gene expression array experiments and the results of pharmacology panels can be queried in the DrugMatrix® database to identify compounds that elicit similar responses or structurally related reference compounds. The third component of DrugMatrix® is an extensive frozen rodent tissue archive that includes snap frozen tissues and corresponding total RNA from liver, heart, kidney, skeletal muscle, whole blood (or plasma) from rats treated with the different test agents.

IV.4.7.2 Key Issue, Hypothesis Tested, or Problem Addressed

A publicly accessible integrated database of rat gene expression profiles, pathology measures, pharmacology assays, and drug literatures profiles on 657 compounds, primarily drugs, will be useful for formulating gene signatures of toxicity, for identifying potentially useful targets for *in vitro* assays, and for linking *in vitro* data to *in vivo* toxicological effects. In addition, the availability of fresh-frozen tissues from the rats used in the toxicological studies will allow for the additional exploration of gene signatures.

IV.4.7.3 Approach

The DrugMatrix® database associates toxicity outcomes based on gene expression data over multiple tissues and times, pharmacology data, pathology data, and relevant published literature. Gene signatures are available for liver, heart, kidney, and skeletal muscle associated with different pathological endpoints. Gene signatures have been formulated for bile duct hyperplasia, hepatic steatosis, necrosis and other pathologies. A general approach for prediction of phenotype would be to upload gene expression data of the 'unknown test agent' into the DrugMatrix® database. The 'unknown substance' gene expression profile would be compared in the database to find compounds with similar expression profile patterns and then analyze the known toxicological effects. DrugMatrix® analyzes the function of significantly altered genes and then enables visualization of 'unknown test substance' effects on 137 toxicological pathways. Genes relevant to a particular pathology or compound class would be created for the 'unknown test substance' being queried. In addition, clustering of gene expression, pharmacology, and blood chemistry data can be performed for the 'unknown test substance' with data from DrugMatrix®.

IV.4.7.4 Results/Progress

Acquisition of the DrugMatrix® database and tissue archive was completed in late September 2010. The database and frozen tissues are being integrated into NTP workflows that include an extensive series of QC tests to evaluate the DrugMatrix® data base and associated tools as described below.

A series of tests have been designed to assess the quality and capabilities of DrugMatrix® upon installation at NIEHS. These include:

- Two hundred Affymetrix.CEL files in the database will be evaluated in GeneSpring for accepted QC metrics.

- All Affymetrix and Unicode files are undergoing QC evaluation.

- Concurrent control/vehicle microarray data are being identified for all chemical treated samples. After DrugMatrix software installation, ten ToxFX reports will be generated by NIEHS and compared with those generated by the contractor.

- RNA will be extracted from 50 random tissue samples for assessment of integrity by Agilent Bioanalyzer for RNA integrity number values >8.

- Functionality and accuracy of the database and corresponding freezer filing system will be evaluated by selecting and validating random selected samples from the tissue repository.

- EPL as the contractor that maintains the tissue archives for the NTP will inspect the DrugMatrix® freezers upon arrival and install liquid nitrogen backup systems.

- A detailed assessment will be made of documentation for the access control mechanisms within the application/system to include detailed descriptions of the database security architecture. The database will be examined for demonstrated public access security controls, data upload architecture, and database security structure that ensures database integrity.

The installation and assessment of DrugMatrix® are currently being performed under the direction of NTP staff. Data in DrugMatrix® will become publically accessible via the NTP CEBS database (http://tools.niehs.nih.gov/cebs3/ui/) in the near future.

IV.4.7.5 Significance

Evaluation of alteration in gene expression is an important component of describing chemically-induced biochemical and pathological changes in tissues. The acquisition of DrugMatrix® is important for many reasons.

1. The database is one of the most comprehensive toxicogenomic compendiums worldwide. Many of the more than 600 test substances are archetypes of their pharmacological or toxicological class of compounds.

2. Toxicogenomic data has already been analyzed in DrugMatrix® to create hundreds of gene signatures phenotypically linked to pathologic, histological, and pharmacologic endpoints. Matching toxicity signatures to gene expression patterns of unknown test substances is a valuable tool for toxicity and target organ prediction.

3. Ready access to relevant toxicological information on DrugMatrix® test compounds facilitates rapid assessment of published literature for unknown test substances under study.

4. The availability of fresh frozen tissues from the rodent tissue repository that are linked to gene expression and toxicological data will be an enormous asset for continued molecular query with advanced molecular technologies for toxicity assessment.

5. Addition of the frozen rodent tissues to the existing NTP archives will greatly enhance its value and overall scope.

IV.4.8 Future Directions/Plans and Justifications

Some of the program objectives in using chemically-induced gene expression changes are to inform selection of *in vitro* assays to help prioritize chemicals for relative toxicity assessment and to elucidate mechanisms of toxicity that will ultimately contribute to improved human risk assessment in regulatory policies and public health. Assay selection for the HTS program undergoes constant improvement from technological advances that result in more informative measures of toxicity. In addition, a continual interplay among data from HTS and *in vitro* and *in vivo* toxicology studies could be mediated by gene expression profiling. If gene expression profiling can derive expression signatures that describe and predict toxic phenotypes of unknown compounds, then it would practical to adapt or model components of these signatures into HTS assays. The purpose of mining NTP tissue archives for gene signatures is to exploit the large collection of tissue block samples and frozen specimens to determine if gene expression signatures can be reliably derived from these samples.

A determination will be made for useful platforms from which to query archival tissues for gene expression changes. Approaches will span those platforms that allow a) in depth genome-wide query into extracted RNA from FFPE tissues such as 3'Seq, to low-throughput but rapid, custom selection of genes such as qPCR, to targeted queries into the archives by mid-throughput platforms such as qNPA.

We are making initial steps toward exploring the archives for molecular signatures. The first goal is to increase our understanding of the quality of extracted RNA from FFPE tissues in the NTP archives. RNA from a number of FFPE tissues, such as liver, kidney and lung, will be extracted to assess quantity, quality, and gene expression of housekeeping and select regulatory genes. Acceptable metrics could be determined for use of an archival study for further molecular profiling. Suitable housekeeping genes (e.g., β-actin, β-2-microglobulin, glyceraldehyde-3-phosphate dehydrogenase, glucuronidase-β, hypoxanthine phosphoribosyltransferase 1, phosphoglycerate kinase 2, protein phosphatase-1α, ribosomal protein L13a, TATA box binding protein, or transferring receptor protein-1) could be evaluated for mouse and rat. After preliminary assessment, three or four housekeeping genes could be selected as surrogates of the transcriptome to evaluate the amplification ability of extracted RNA in the testing of archival specimens. Evaluation of a representative set of tissue from two or three studies could be accomplished using qNPA analysis or by custom qPCR arrays. These studies could be expanded to include a desired number of tissue and organ types for assessment, noting any differences among them in expression of housekeeping genes. Also, the utility of extractable RNA from FFPE tissues stored for different lengths of time need to be assessed. The aging of tissue blocks from increasing older studies would be assessed to provide data on the relative usefulness of studies completed within 2-5 years, 5-10 years, and 10-20 years. A comparison among tissue types, gender, species and any additional factors could be measured for usefulness in molecular characterization for FFPE specimens. This same approach could be performed on FFPE tissue from the NTP archives for extracted DNA for epigenetic changes. Methylated DNA has been reported to be stable in FFPE tissues. An evaluation of the ability to identify methylated sites within DNA for correlation with gene expression changes could be approached in a similar manner for different organs and tissues, over time, gender, and species as described above.

As a next step, these methods could be applied in discovery mode into the transcriptome using 3'Seq for wide coverage for compounds of known organ or tissue specific toxicity. Gene signatures could be derived for target tissues and non-target tissues and then those gene signatures could be screened in the NTP tissues archives by qNPA analysis depending upon the size of the signature and level of throughput desired. For example, custom array plates can be designed in the following formats: 96-well plate with up to 47 genes per well measured or a 384 well plate with up to nine genes per well

measured. Similarly, epigenetic assessments could also be made upon archival tissues of interest. High throughput assays might be designed or sought out for those genes in a signature that are believed to be most representative of a toxicological or mechanistic process. The potential for gene signatures in DrugMatrix® signatures to impact HTS assays and expression analysis of NTP archival tissues will soon be explored as it becomes part of the NTP infrastructure.

IV.4.9 References

Auerbach SS, Shah RR, Mav D, Smith CS, Walker NJ, Vallant MK, et al. 2010. Predicting the hepatocarcinogenic potential of alkenylbenzene flavoring agents using toxicogenomics and machine learning. Toxicol Appl Pharmacol 243(3):300-14.

Achanzar WE, Diwan BA, Liu J, Quader ST, Webber MM, Waalkes MP. 2001. Cadmium-induced malignant transformation of human prostate epithelial cells. Cancer Res 61:455-458.

Achanzar WE, Brambilia EM, Diwan BA, Webber MM, Waalkes MP. 2002. Inorganic arsenite-induced malignant transformation of human prostate epithelial cells. J Natl Cancer Inst 94:1888-1891.

Beck AH, Weng Z, Witten DM, Zhu S, Foley JW, Lacroute P, et al. 2010. 3'-End sequencing for expression quantification (3SEQ) from archival tumor samples. Plos One 5(1):e8768.

Klaassen CD, Aleksunes LM. 2010. Xenobiotic, bile acid, and cholesterol transporters: function and regulation. Pharmacol Rev 62(3):1-96.

Iliopoulos D, Guler G, Han SY, Druck T, Ottey M, McCorkell KA, et al. 2006. Roles of FHIT and WWOX fragile genes in cancer. Cancer Lett 232(1):27-36.

Jonker JW, Schinkel AH. 2004. Pharmacological and physiological functions of the polyspecific organic cation transporters: OCT1, 2, and 3 (SLC22A1-3). J Pharmacol Exp Ther 308(1):2-9.

Körner M, Miller LJ. 2009. Alternative splicing of pre-mRNA in cancer: focus on G protein-coupled peptide hormone receptors. Am J Pathol 175:461-472.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25.

Mardis ER. 2008. Next-generation sequencing methods. Ann Rev Genomics Hum Genet 9:387-402.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621-628.

Murphy G. 2008. The ADAMs: signaling scissors in the tumour microenvironment. Nat Rev Cancer 8:929-941.

Notarbartolo M, Cervello M, Giannitrapani L, Meli M, Poma P, Dusonchet L, et al. 2004. Expression of IAPs and alternative splice variants in hepatocellular carcinoma tissues and cells. Ann NY Acad Sci 1028:289-293.

Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. 1999. Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinog 24(3):153-159.

Sala A, Watson R. 1999. B-Myb protein in cellular proliferation, transcription control, and cancer: latest developments. J Cell Physiol 179(3):245-50.

Sampath J, Pelus LM. 2007. Alternative splice variants of survivin as potential targets in cancer. Curr Drug Discov Technol 4(3):174-91.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Res 19(6):117-1123.

Sze KM, Ching YP, Jin DY, Ng IO. 2008. Role of novel splice variant of mitotic arrest deficient 1 (MAD1), MAD1beta, in mitotic checkpoint control in liver cancer. Cancer Res 68(22):9194-9201.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics 25(9):1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511-515.

IV.5 Tox21 Activities:

A Bioinformatics-Based Approach to Identifying Assays That Query Human Health Effects

This page left intentionally blank

IV.5 Tox21 Activities: A Bioinformatics-Based Approach to Identifying Assays That Query Human Health Effects

Scott Auerbach, Ph.D., DABT
Biomolecular Screening Branch

Associated NTP Scientists
Kristine Thayer, Ph.D. (CERHR)
Vickie Walker (CERHR)

Collaborating Scientists
Richard Judson, Ph.D. (U.S. EPA NCCT)
Allison Mandich (NHGRI/NCGC)

IV.5.1 Background and Rationale

Some of the most common causes of morbidity and mortality in the developed world, such as diabetes, obesity, heart disease, and neurological disease, are not comprehensively queried in standard chemical toxicological studies. Often expensive, specifically designed studies are needed to capture chemical hazards related to these diseases. For this reason, *in vitro* assays that can provide signals to justify such studies would be of significant value.

As genomics has advanced, the scientific community has begun to understand the molecular events that lead to a wide-variety of human health effects. Such discoveries have allowed for the *in vitro* modeling of disease-related pathophysiological processes; models that can be used to determine the impact of environmental agents on such processes. Before a large scale endeavor can be undertaken to model and query human disease processes *in vitro*, the relationships between disease biology and testable biological space needs to be established.

IV.5.2 Key Issue, Hypothesis Tested, or Problem Addressed

The goal of this project is to create meta-database that relates genes, pathways, and biological processes to human disease and subsequently to identify the chemical genomic space within these relationships that can be exploited to query the effects of chemicals on molecular processes related to human disease.

IV.5.3 Approach

The first step is to merge data from a large of disease/genome databases, including the following:

- The Comparative Toxicogenomics Database (http://ctd.mdibl.org/)(Davis et al. 2010)
- Phenopedia (http://www.hugenavigator.net/HuGENavigator/startPagePhenoPedia.do)(Yu et al. 2009)
- Human Genome Association Database (http://geneticassociationdb.nih.gov/)(Zhang et al. 2010)

- OMIM (http://www.ncbi.nlm.nih.gov/omim)(Sayers et al. 2009)
- GeneCards (http://www.genecards.org/)(Safran et al. 2010)
- Entrez Gene (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene)(Sayers et al. 2009)
- CoPub (http://services.nbic.nl/copub/portal/)(Frijters et al. 2008)
- KEGG Disease (http://www.genome.jp/kegg/disease/)(Kanehisa et al. 2009)

These resources procure information from a number of resources including literature mining, genetic studies, and functional genomics studies. Disease:gene relationships found in these databases are cataloged and a weighted voting approach is then used to identify a rank list of genes for each disease of concern. It is then determined if the identified disease genes possess a protein domain that would provide an interaction interface for an environmental agent. Disease genes that fit this category become a priority for assay identification/generation. In addition to the gene-centric approach to assay selection, pathway and biological process level approaches to assay will also be implemented. Disease:pathway/biological process relationships will be determined by the enrichment of genes in the identified disease categories. Enriched pathway/biological process categories will serve as a basis for selection of additional assays to query chemical hazard as it relates to specific diseases.

The described approach in large part depends on literature mining and therefore publication bias can potentially inflate the rank of a gene. In addition, understudied/under published disease:gene relationships may lead to a deflated rank for potentially important genes.

IV.5.4 Results/Progress

An analysis of type 2 diabetes and obesity has identified a number of target genes that can serve as the basis for assays to screen for chemicals that may perturb these human health effects (**Tables IV.5-1** and **IV.5-2**).

Table IV.5-1 Top 10 genes identified for Diabetes mellitus, type 2, assay development

Gene Symbol	Official Gene Name
PPARG	peroxisome proliferator-activated receptor gamma
KCNJ11	potassium inwardly-rectifying channel, subfamily J, member 11
HNF4A	hepatocyte nuclear factor 4, alpha
ABCC8	ATP-binding cassette, sub-family C (CFTR/MRP), member 8
ENPP1	ectonucleotide pyrophosphatase/phosphodiesterase 1
GCGR	glucagon receptor
PPARA	peroxisome proliferator-activated receptor alpha
LPL	lipoprotein lipase
APOB	apolipoprotein B (including Ag(x) antigen)
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1
SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1

Table IV.5-2 Top 10 genes identified for obesity assay development

Gene Symbol	Official Gene Name
PPARG	peroxisome proliferator-activated receptor gamma
PPARA	peroxisome proliferator-activated receptor alpha
SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1
PPARD	peroxisome proliferator-activated receptor delta
INSR	insulin receptor
ADRB3	adrenergic, beta-3-, receptor
PCSK1	proprotein convertase subtilisin/kexin type 1
ENPP1	ectonucleotide pyrophosphatase/phosphodiesterase 1
CNR1	cannabinoid receptor 1 (brain)
NPY	neuropeptide Y

IV.5.5 Significance

The database developed with this effort provides a disease focus to the assay development and prioritization process. In addition, the assays identified using the described approach will have the ability to detect perturbations of critical pathways in human disease. With this knowledge, more refined and targeted testing of greater human disease relevance may be possible. Eventually, it may be possible to use the assays identified using the described approach for human health risk assessment, therefore limiting the use of animals for specialized hazard characterization.

IV.5.6 Future Directions

In the initial stages, a gene-centric approach has been taken which focuses on binary relationships between a single gene and a disease. The typical result of such an approach is the identification of genes that are amenable to pharmacology assays. Such assays have their limitations with respect to cellular and biological processes. For this reason, future work will emphasize the relationship between pathways/biological processes and disease. Establishment of these relationships will allow for identification of cellular phenotype/pathway level assay development, which will more closely approximate the integrated molecular events in toxicity.

Non-reactive xenobiotics, particularly non-pharmaceuticals, elicit their toxicological effects through binding to receptors and enzymes that possess promiscuous binding sites (e.g., PXR, CYP3A4) (Nobeli et al. 2009). Through an automated, in silico, structure-based analysis we hope to identify those targets encoded in the genome that display a high degree of promiscuity (high promiscuity index). Promiscuity index values derived from this analysis will be integrated into our target selection process with the goal of identifying targets that are most likely to be effected by environmental chemicals.

IV.5.7 References

Davis AP, King BL, Mockus S, Murphy CG, Saraceni-Richards C, Rosenstein M, et al. 2010. The Comparative Toxicogenomics Database: update 2011. Nucleic Acids Res Sep 22. [Epub ahead of print]

Frijters R, Heupers B, van Beek P, Bouwhuis M, van Schaik R, de Vlieg J, et al. 2008. CoPub: a literature-based keyword enrichment tool for microarray data analysis. Nucleic Acids Res 36(Web Server issue): W406-410.

Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. 2009. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38(Database issue): D355-360.

Nobeli I, Favia AD, Thornton JM. 2009. Protein promiscuity and its implications for biotechnology. Nat Biotechnol 27(2): 157-167.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. 2009. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 38(Database issue): D5-16.

Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. 2010. GeneCards Version 3: the human gene integrator. Database (Oxford) 2010: baq020.

Yu W, Clyne M, Khoury MJ, Gwinn M. 2009. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. Bioinformatics 26(1): 145-146.

Zhang Y, De S, Garner JR, Smith K, Wang SA, Becker KG. 2010. Systematic analysis, comparison, and integration of disease based human genetic association data and mouse genetic phenotypic information. BMC Med Genomics 3: 1.

IV.6 Tox21 Activities:

The Mouse Methylome Project

This page left intentionally blank

IV.6 Tox21 Activities: The Mouse Methylome Project

John E. (Jef) French, Ph.D.
Biomolecular Screening Branch

Biomolecular Screening Branch
Scott S. Auerbach, Ph.D., DABT
B. Alex Merrick, Ph.D.

Associated NTP Scientists
Angela King-Herbert, DVM, Ph.D. (CMPB)
Dave Malarkey, DVM, Ph.D. (CMPB)

Associated NIEHS Scientists
Trevor Archer, Ph.D. (LMC)
Takashi Shimbo, Ph.D. (LMC)
Keith Shockley, Ph.D. (BB)
Paul Wade, Ph.D. (LMC)

Collaborating Scientists
Simon Gregory, Ph.D. (Duke University, Durham, NC)
Jim Mullikin (NHGRI, NIH Intramural Sequencing Core, Rockville, MD)
Ruchir Shah, Ph.D. (SRA International)

IV.6.1 Background and Rationale

An individual's response to exposure related toxicity and concomitant disease is influenced at the genome level by genetic, epigenetic, gene-gene interactions (**intrinsic factors**), and interaction with the environment (**extrinsic factors**). Individual DNA sequence variation does not account for all of the heritability for susceptibility to toxicity and diseases such as asthma, cancer, or diabetes. One intrinsic factor that quantitative and molecular geneticists believe could contribute to the observed "missing heritability" is the **methylome** (Eichler et al. 2010), an individual's genome wide pattern of cytosine methylation. The methylome (a component of the epigenome) may be a major epigenetic modifier of the susceptibility to cancer and other chemical exposure related diseases. Major basic questions about the nature of epigenetic variation within individuals remain understudied. It is not known whether, or to what extent, DNA methylation patterns (or other epigenetic marks) are inherited from parent to offspring. Likewise, variability in patterns of DNA methylation within individuals is not well described at the genome level. Although extrinsic factors are hypothesized to impact the epigenome (and the methylome), the extent to which such interactions influence biological outcomes of exposure, and whether any such effects are heritable, remain unclear.

Presently, there is no mouse reference database for the methylome akin to the NTP/Perlegen DNA sequence database of 15 commonly used inbred strains (plus the C57BL/6 reference strain)(Frazer et al. 2007, Yang et al. 2007). The DNA sequence data has significantly increased our knowledge of the genomic structure of the inbred mouse and has provided the basis for imputation of the haplotype structure of more than 90 inbred strains used in biological research (Kirby et al. 2010). The absence of a

methylome reference database for the mouse significantly handicaps our knowledge and understanding of the mouse model in toxicology and environmentally related diseases and hinders the design and performance of hypothesis based genetic and epigenetic research studies to understand the associated mechanisms and relationships.

Two high content technologies have been recently developed that 1) permit genome-wide determination of cytosine methylation, DNA sequence variation, and RNA sequence at base pair resolution (massively parallel sequencing, bisulfite seq (BIS seq), RNA seq) from a single biological sample; and 2) fractionate DNA sequences using differential restriction and/or affinity capture (MMDE-seq) to enrich for methylated DNA sequences from limited quantities of biological material including FFPE tissue samples (Bormann Chung et al. 2010, Down et al. 2008, He et al. 2010, Hughes and Jones, 2007, Jacinto et al. 2008, Mill and Petronis, 2009, Mill et al. 2006, Serre et al. 2009). Together, these tools allow targeted interrogation of genomic regions of interest using bioinformatic data mining tools. The proposed study will use these technologies to create a definitive map of the mouse liver methylome from the two parental strains (C57BL/6N and C3H/HeN) and their F1 hybrid (B6C3F1/N) offspring that exhibit dramatically different rates of intrastrain and interstrain as well as sex dependent spontaneous liver cancers. The high, but variable, incidence of liver tumors in the F1 hybrid mouse often confounds interpretation of 2-year toxicology and carcinogenesis studies. Although highly penetrant quantitative trait loci have been identified in the C3H/He strain, the liver cancer incidence varies significantly in untreated control B6C3F1 mice from generation to generation. This variable incidence may be due in part to cytosine methylation in critical tumor suppressor genes, regulatory regions of the genome, and associated pathways. The reference database will aid our understanding of the relationship between variations in sporadic and induced disease incidence associated with individual variations in the methylome, DNA sequence, and exon specific transcript expression critical to understanding the potential functional consequences from generation to generation. The data from this project, will directly address critical knowledge gaps in the nature of the methylome, its variability across individuals, its association with disease, and its heritability across generations. Further, these data will create a reference for future investigations into environment-induced changes in methylome variation and its role in spontaneous and induced disease.

IV.6.2 Key Issue, Hypothesis Tested, or Problem Addressed

Scientific evidence indicates that individual differences in susceptibility to toxicity, sporadic disease, and exposure related disease within a population of individuals is based upon individual differences in genomic structure (e.g., individual differences in SNPs, and copy number variation [CNV]). However, this variation explains only part of the heritability of susceptibility to disease. The proposed research will facilitate our understanding of the role of individual differences in epigenomic structure within and between individuals, heritability of DNA methylation patterns, and the relationship of these events to disease susceptibility. By determining an individual's genomic DNA sequence, methylated sequences, and the exon-specific transcript expression, we can correlate genome wide methylation patterns with tissue specific transcript expression and create a reference database under controlled conditions. The mouse has been shown to be an excellent model organism for many human diseases. Using syngeneic individuals from a population of inbred mice analogous to studies of human monozygotic twins, we can work toward separating "signal from noise" by understanding individual variations in SNPs, CNV, and cytosine methylation patterns that may affect micro RNA, long non-coding RNA and messenger RNA expression. These efforts are expected to facilitate identification of target sequences for hypothesis-based research in environmental health.

IV.6.3 Approach (Research Plan)

IV.6.3.1 Aims

1. Sequence and catalog the genomic DNA sequence, genomic cytosine methylation pattern, and exon specific transcripts of both sexes of the C57BL/6N (B6) strain and C3H/HeN (C3) strain and the female B6 x male C3 and the female C3 x male B6 outcrosses that produces the female and males B6C3F1/N or C3B6F1/N hybrid progeny, respectively, to create a reference mouse DNA genome, methylome, and transcriptome database for each sex of each inbred strain under standard NTP specifications and controlled study and environmental conditions.

2. Determine the individual intra- and inter-strain differences in cytosine methylation (B6 and C3, and the F1 hybrid female and male) that may potentially explain differential toxicity and tissue specific disease outcomes within and between these inbred strains and their F1 hybrid.

3. Determine, correlate, and catalog each strain's individual methylome with its exon specific transcriptome (microRNA, long non-coding RNA, and messenger RNA transcripts) at both the quantitative (expression level) and qualitative (splicing) level.

4. Determine, correlate, and catalog heritable regions of the methylome (DNA sequence specific cytosine methylated sequences) and the heritability of the transcriptome of each strain.

IV.6.3.2 Study Design

Animals: C57BL/6N (B6) female and male, C3H/HeN (C3) female and male, and their B6C3F1/N and C3B6F1/N female and male hybrid progeny will be used (Source: NTP Colony). All breeders and offspring mice will be clearly identified by tattoo and lineage. All mice will have specified tissues sampled at the same age and after being raised under the same environmental conditions as described.

Breeding Scheme: A total of 10 female and 10 male B6 mice and 10 male and 10 female C3 mice will be randomly selected from the NTP strain maintenance colony. One pair of female and male siblings from the same litter will be randomly selected from each breeding unit until 10 of each sex-strain pair have been isolated and uniquely identified after weaning and sex has been confirmed. From this population of randomly selected mice of each strain, 7 breeding pairs of B6 females and C3 males and 3 breeding pairs of C3 females and B6 males will be randomly paired for outcross to produce B6C3F1 or C3B6F1 hybrid female and male progeny. All mice will be uniquely identified and their lineage tracked and confirmed.

At 10 weeks of age, female mice and male mice will be randomly selected and pair mated to carry out the conventional and the reciprocal backcross as described. After confirmation of pregnancy, the male mice will be removed and housed separately. At weaning, the B6C3F1/N and C3B6F1/N hybrid pups will be uniquely identified, sexed, and the females and males housed separately and their lineage tracked. At 17 weeks of age, their diet will be changed from NIH31 to NTP 2000 until each has reached 20 weeks of age.

Total mice: **Parental lines**: 10 mice/sex x 2 sexes x 2 strains (B6 and C3) = 40 mice to be used for tissue sampling as specified. **Progeny**: 1 mouse/sex x 2 sexes x 10 breeding pairs = 20 mice plus 2 additional female and male siblings from one randomly selected paired mating for both B6C3F1/N and C3B6F1/N (2 sexes x 2 siblings x 2 outcrosses = 8 mice) to be used for tissue sampling as specified. TOTAL: 68 mice

(10 male and 10 females/strain plus 3 siblings/sex from 2 litters (B6C3F1/N and C3B6F1/N) for sampling and tissue prepared for sequencing and archival.

Environment: Conducted under standard NTP Specifications without test agent exposure.

Diet: All mice will be placed on NIH31 (NTP breeding diet) for the first 17 weeks of life and then switched to NTP2000 (study diet) for the final 3-week prior to euthanasia and tissue collection. This scheme is the best approximation of diet and dietary exposures in the production of B6C3F1/N hybrids for NTP studies.

Tissue samples: The primary tissue of initial interest is the liver. Liver was selected for the primary analysis because of its relative homogeneity (80-90% hepatocytes) and its relative importance to NTP carcinogenicity studies in the B6C3F1/N mouse (as noted above). The left lateral lobe will be rapidly removed and dissected in a dish on ice into 3 – 4 mm cubes and flash frozen in 1 mL Eppendorf screw cap tubes in liquid nitrogen and stored at -80°C prior to sample preparation for DNA isolation and library production for bisulfite sequencing and exon specific transcript expression analysis of individual mouse genomes.

Other tissues (adipose, brain, cardiac muscle, and skeletal muscle) will also be collected as described by multiple prosectors as rapidly as possible, flash frozen, and archived for future studies as warranted.

Molecular studies: Up to 3 liver samples for each sex-strain pair will be initially investigated to determine the within and between strain variation. More samples may be analyzed as required after statistical evaluation of the results

Phase 1 – C57BL/6N female, C3H/HeN male, B6C6F1 male and female

Phase 2 – C3H/HeN female, C57BL/6N male, C3B6F1 male and female

Phase 3 – additional replicates and targeted resequencing as necessary

We will incorporate into our study:

Bisulfite sequencing (BIS-Seq) with DNA sequence genomic controls

Targeted re-sequencing of specific sites (MMDE-seq)

Whole exon-specific transcriptome expression profiles (microRNA, long non-coding RNA, and messenger RNA transcripts; RNA-Seq)

IV.6.3.3 Data Analysis

1. Determine and catalog the genome wide cytosine methylation patterns: Bioinformatic methods will be employed to align and map reads to the genome of the 3 strains to create a high resolution map of the methylome

2. Determine intra-strain variation by sex: Identify differentially methylated sites within females and males of both parental strains (B6 and C3) and their F1 hybrid

3. Determine inter-strain variation by sex: Identify differentially methylated sites between the parental strains (B6 and C3) and their F1 hybrid

4. Determine whether sexual dimorphisms in DNA methylation patterns exist, and their extent across the genome.

5. Identify association between local methylation patterns in the genome and both quantitative and qualitative variation in the transcriptome. Correlative analysis will be performed to identify these relationships.

6. Identify heritable regions of the methylome: Compare the methylome of each of the parental B6 females to their B6C3F1/N female offspring and the C3 males to their B6C3F1/N male's offspring. For the reciprocal outcross, compare the methylome of each of the parental C3 females to their C3B6F1/N female offspring and the B6 males to their C3B6F1/N male offspring to determine effects of germline transmission of imprinted genes (cytosine methylation variation in known imprinted genes).

Liver samples from all mice at 20 weeks of age will be collected and flash frozen as described above in "Tissue Sample". From each of the 10 breeding pairs, the female C57BL/6N dam along with a male sibling and the male C3H/HeN sire along with a sibling female will have liver samples collected (10 pairs x 2 sexes x 2 strains = 40 samples). From the F1 progeny, 1 randomly selected male and 1 randomly selected female B6C3F1/N from each set of the 10 mating pairs will have liver samples collected and frozen as described. In addition, one set of B6C3F1 sibling females and males will be randomly identified and 3 individuals of each sex will also have liver samples collected [(10 pairs progeny x 1 strain x 2 sexes) + (2 progeny x 2 sex x 2 strains)] = 28 samples. Thus, there will be a total of 68 liver sample sets will be collected and available for processing and for analysis.

In the first phase of sequencing, only the liver samples from one B6 female (dam), one C3 males (sire), one male and one female B6C3F1/N hybrid (2 genomes, 4 BIS-Seq and RNA-Seq total) of the conventional outcross will be sequenced and analyzed. In the second phase, only the liver samples from one C3 female (dam), one B6 male (sire), and one female and one male C3B6F1/N hybrid (2 genomes, 4 BIS-Seq, and 4 RNA-Seq total) of the reciprocal outcross will be sequenced and analyzed. This will allow sufficient sequencing data to be analyzed and processed to determine the strategy for the third phase of sequencing to answer the question of within and between strain and within litter variation relative to the nature of the outcross and germline effects relevant to transgenerational outcomes. Further analysis (increasing the number of observations per strain/sex) samples will allow examination of the within and between strain variation of the males (C3 and B6C3F1) and the females (B6 and B6C3F1) and the reciprocal cross (if warranted). Further sequencing of samples will be carried out as warranted to complete the study aims by permitting comparison of sexual differences within and across the two inbred strains (B6 and C3) and by providing sufficient replicates to discriminate "signal from noise" (i.e., random vs. non-random cytosine methylation). Estimation of the strain and sex variation in stochastic methylation is deemed critical to identification of candidate genomic regions for hypothesis-based research on the epigenetic basis of strain and sexual variation in spontaneous liver cancer incidence.

IV.6.3.4 Tiered Approach: Number of mice to be sequenced and analyzed

The initial critical question being asked in the first phase of this project is: What is the extent and genomic location of cytosine methylation variation within the B6 female and male and the C3 female and male inbred strains and within female and male siblings of a heterozygous hybrid strain of the conventional and reciprocal outcross. By defining this variation in methylated cytosine sequences and the corresponding DNA sequence context genome-wide, targeted sequencing strategies can be developed for hypothesis-based research, including the basis for trans-generational effects. To control

efficiency and cost, three individuals (n = 2 or 3) of each sex and genotype will be considered the minimum number necessary to discriminate between random and non-random variation in methylated CpG sequences and the maximum number in regard to sequencing and development of high content data set costs. After examining the initial sequence results, sequencing of additional samples, including female and male B6 and female and male C3 to estimate within sex and strain variation and to increase the number of observation to more than 3 samples each sex/strain may be required in order to increase the power of detection.

IV.6.4 Results/Progress

Breeding, sample collection, and archival of tissues will be undertaken by NTP and is in progress. Libraries for whole genome and bisulfite sequencing will prepared by the laboratory of Paul Wade, LMC, DIR, who have significant experience in the preparation of libraries. Bioinformatic analysis will be carried out by NIEHS and NIH bioinformaticians. The analysis tools for massively parallel sequencing or NextGen sequencing are still being developed and a major bioinformatic effort will be required. This effort will include sequence alignment and computational analysis to correlate DNA sequence, methylated sequence, and exon-specific transcripts.

IV.6.5 Significance

The genetic basis (SNPs, CNV, somatic mutations, etc.) for susceptibility explains only part of the role of individual variation in heritable and derived phenotypic traits, including sporadic and environmental related diseases such as asthma, cancer, diabetes, obesity, etc. Inbred mouse models share significant features in genetic and genomic structures and susceptibility to disease with humans. The range of genetic variation in laboratory and wild-derived strains of mice is similar in magnitude to the variation in SNPs and CNV observed in human populations. This project will address fundamental questions in regard to DNA sequence and the methylome, its variability, heritability, relationship with gene expression, and with disease. Further, this project will propel construction of a reference database for the methylome of inbred strains as a research tool available to the NTP, DIR, and the broader scientific community. It is anticipated that such information may spur further studies investigating the mechanistic basis by which the epigenome intersects with environmental exposure in disease incidence, susceptibility and severity.

IV.6.6 Future Directions/Plans and Justifications

If successful and warranted by the results, we plan to extend the approach to other tissues collected from these animals and to additional inbred strains sequenced previously in the NTP-Perlegen research project. The intent is to develop a cohort of genetically and epigenetically diverse set of inbred strains, characterized across several tissues of interest for quantitative analysis of toxicity and disease phenotypes through haplotype and/or meiotic association mapping.

Targeted differential restriction and high throughput bisulfite sequence along with the DNA sequence can be used to develop genetic and epigenetic marker reference data set for high-resolution genome wide haplotype association mapping. Those highly significant markers within or near reference methylated cytosine sequences that are associated with quantitative traits of interest, targeted genome wide HTS bisulfite sequencing can be used for high resolution mapping and the data used to confirm SNP and CpG markers for haplotype association mapping of phenotypic traits. Highly significant candidate sequences may be further examined for functional validation using *in vitro* or *in vivo* models for single and multiple (trans-) generation studies.

Develop the tools to determine the value of using the FFPE archived tissues and targeted HTS bisulfite sequencing to investigate mechanisms of toxicity and disease association within NTP studies, and, where possible, compare NTP liver cancer specimens with DIR human liver specimens to address the relevance to human disease.

IV.6.7 References

Bormann Chung CA, Boyd VL, McKernan KJ, Fu Y, Monighetti C, Peckham HE, et al. 2010. Whole methylome analysis by ultra-deep sequencing using two-base encoding. PLoS One 5(2): e9320.

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, et al. 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 26(7): 779-785.

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 11(6): 446-450.

Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ, et al. 2007. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature 448(7157): 1050-1053.

He D, Choi A, Pipatsrisawat K, Darwiche A, Eskin E. 2010. Optimal algorithms for haplotype assembly from whole-genome sequence data. Bioinformatics 26(12): i183-190.

Hughes S, Jones JL. 2007. The use of multiple displacement amplified DNA as a control for methylation specific PCR, pyrosequencing, bisulfite sequencing and methylation-sensitive restriction enzyme PCR. BMC Mol Biol 8: 91.

Jacinto FV, Ballestar E, Esteller M. 2008. Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. Biotechniques 44(1): 35, 37, 39 passim.

Kirby A, Kang HM, Wade CM, Cotsapas C, Kostem E, Han B, et al. 2010. Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. Genetics 185(3): 1081-1095.

Mill J, Petronis A. 2009. Profiling DNA methylation from small amounts of genomic DNA starting material: efficient sodium bisulfite conversion and subsequent whole-genome amplification. Methods Mol Biol 507: 371-381.

Mill J, Yazdanpanah S, Guckel E, Ziegler S, Kaminsky Z, Petronis A. 2006. Whole genome amplification of sodium bisulfite-treated DNA allows the accurate estimate of methylated cytosine density in limited DNA resources. Biotechniques 41(5): 603-607.

Serre D, Lee BH, Ting AH. 2009. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res 38(2): 391-399.

Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F. 2007. On the subspecific origin of the laboratory mouse. Nat Genet 39(9): 1100-1107.

This page left intentionally blank

V. Future of Tox21 at NTP

This page left intentionally blank

V. Future of Tox21 at NTP

Raymond R. Tice, Ph.D.
Chief, Biomolecular Screening Branch

Tox21 is a collaborative effort initiated originally among three Federal organizations with diverse experiences and skills, including the experimental toxicology expertise at the NTP, the qHTS technology of the NCGC, and the computational toxicology capabilities of the NCCT, to change the face of toxicology in the 21st Century. Our collaboration began in 2006 prior to the 2007 publication of the NAS report on "Toxicity Testing in the 21st Century: A Vision and Strategy" (http://www.nap.edu/catalog.php?record_id=11970) but was formalized in response to the NAS report with the November 14, 2008 release of a MOU on "High Throughput Screening, Toxicity Pathway Profiling, and Biological Interpretation of Findings". The recent addition of the U.S. FDA increases the breadth of experience in human diseases and in animal models of human disease, as well as in toxicity pathway analysis and computational toxicology.

The central component of Tox21 is the exploration of HTS assays and tests using phylogenetically lower animal species (e.g., zebrafish embryos, *C. elegans*), as well as high throughput, whole genome analytical methods to evaluate mechanisms of toxicity. Ultimately, the data generated by these new tools are to be provided to the scientific and regulatory community to use in the protection of human health and the environment. As presented earlier, the goals of Tox21 are to use these tools to:

- identify mechanisms of chemically induced biological activity
- prioritize chemicals for more extensive toxicological evaluation
- develop more predictive models of *in vivo* biological response, with a focus on humans

Specifically, the related goals of the BSB are to:

- carry out the NTP's automated screening assays with *C. elegans*
- develop research and testing activities in high and medium throughput screening assays for rapid detection of biological activities of significance to toxicology and carcinogenesis
- develop computational tools and approaches to allow an integrated assessment of HTS endpoints and associations with findings from traditional toxicology and cancer models
- develop assays and approaches to understand the genetic and epigenetic bases for differences in susceptibility

This report and the platform and poster presentations during the BSC meeting provide information on the:

- history, purpose, and structure of the Tox21 Initiative
- capabilities and scope of activities that each partner brings to Tox21
- past, current, and projected future activities of each Tox21 Working Group

- major activities initiated by the BSB in order to potentially achieve the vision of the NRC report and the goals of Tox21

The presentations on our main Tox21 activities focused on studies:

- conducted by the NTP *C. elegans* "Worm Tox" core laboratory to develop and evaluate toxicological assays using this nematode.

- initiated in-house or via an extramural collaboration to probe mechanisms of inter-individual susceptibility to toxicants using cells from densely genotyped humans and strains of mice in qHTS.

- to evaluate the application of recent advances in molecular biology for obtaining signature. expression profiles from FFPE tissues in the extensive NTP tissue archives.

- to create a meta-database that relates genes, pathways, and biological processes to human disease, and subsequently to identify the chemical genomic space within these relationships that can be exploited to query the effects of chemicals on molecular processes related to human disease. This approach was used by BSB staff to suggest pathways that might be involved in autism at an NIEHS-sponsored workshop on this disease held in September 2010, and is being used to provide pathway-based information on diabetes and obesity for a workshop on "State-of-the Science Evaluation of Environmental Exposures and Diabetes/Obesity", sponsored by the NTP Center for the Evaluation of Risks to Human Reproduction (CERHR), to be held on January 11-13, 2011, at NIEHS.

- to determine the methylome of inbred strains of mice in order to create a reference for future investigations into environment-induced changes in methylome variation and its role in spontaneous and induced disease.

The BSB staff is involved in other Tox21-related projects as well; some of which include:

- developing SBIR contracts awarded in 2010 to (1) develop mid- to high-throughput toxicological tests using model organisms (i.e., zebrafish embryos), (2) develop quantitative high throughput screens for the detection of chemicals that modulate gap junction intercellular communication, (3) incorporate metabolism into quantitative high throughput screening assays, (4) develop an integrated prediction systems to support environmental toxicological assessments, and (5) produce chemical/stress responsive transgenic C. elegans in order to monitor *in vivo* gene expression changes after exposure to toxicants.

- developing SBIR contracts to be awarded in 2011 that focus on the (1) development of a new high throughput screening for reactive oxygen species mediating toxicity, (2) development of *in vitro* 3D tissue models (e.g., lung, kidney, skin) for toxicity testing (an important bridge between *in vitro* monocultures and *in vivo* testing), and (3) application of new 'omics technologies to rodent FFPE tissue samples, to support the development of methods and tools that enable the use of FFPA tissues for next-generation sequencing analysis of the genome, transcriptome, and epigenome.

- efforts within NTP by members of the BSB and the Toxicology Branch to develop the models needed to extrapolate from in vitro concentration-based data to in vivo administered dose

and/or resulting blood concentration levels, taking into account absorption, distribution, metabolism, and excretion kinetics.

- collaborations with NIEHS intramural scientists (e.g., Drs. Perry Blackshear, Anton Jetten, Samuel Wilson) to screen the NTP compound library against targets of potential interest from both a pharmaceutical and a toxicological viewpoint at the NCGC in qHTS; as well as a metabolomics study with Dr. Jetten using knock out mice for RORα (retinoid-related orphan receptor alpha) to potentially establish a link between interaction of xenobiotics with this gene, its effects on gene expression, physiological processes (including metabolism), and disease.

- collaborations with NIEHS extramural scientists that have included providing the NTP 1408 compound library for use in screening against a target of environmental interest (e.g., Dr. Eileen Jaffe, Fox Chase Cancer Center, Philadelphia, PA); providing samples from human lymphoblastoid cells from the NCGC for an analysis of epigenetic changes caused by environmental compounds (Dr. Art Petronis, the Krembil Family Epigenetics Laboratory Centre for Addiction and Mental Health, Toronto, Canada), comparing the results from docking models for the nuclear receptors screened in Phase I at the NCGC with the experimentally obtained data on the NTP and EPA compound libraries (Dr. Michael Goldsmith, U.S. EPA NHEERL), and comparing the predictions of a toxiphore model for mitochondria toxicity developed by Dr. Craig Beeson (Medical University of South Carolina, Charleston, SC) with mitochondria toxicity data obtained on the NTP compound library in qHTS at the NCGC.

The data generated by Tox21 are being used or will be used by NTP to help:

- identify/prioritize compounds for more extensive toxicological testing.

- rank compounds within the same class with regard to potential hazard.

- categorize multiple forms of a complex mixture (e.g., herbal products) into different “biological activity bins”, based on patterns of response across multiple mid- and high-throughput screens.

- identify the kinds of animal studies that should be conducted and/or endpoints that should be evaluated in these studies, based on the pattern of activity of a compound across different mid- and high-throughput screens.

- interpret results obtained in classical toxicological studies.

- assess the differential response of cells from various inbred strains of mice to interpret the relationship between intrinsic genetics, chemical sensitivity, and disease.

- identify key cellular pathways linked to disease and the environmental compounds that might contribute to the appearance of that disease.

- develop prediction models for rodent and human disease, the former because of the extensive rodent data that exists and because the validity of a prediction model for disease is most easily tested in laboratory rodents.

The scope of Tox21 in terms of (1) the numbers and chemical coverage of compounds being screened for activity across a broad spectrum of *in vitro* and alternative model organism based assays of toxicological interest, and (2) the databases, knowledgebases, and computational/informatic tools being

made publicly available has resulted in considerable interest among the international scientific and animal welfare communities in our efforts and of collaborating in our activities. Recent examples include:

- several pharmaceutical companies have provided samples of potential drugs that failed during clinical trials to the U.S. EPA NCCT for screening in ToxCast™ Phase II and at the NCGC; structures and all data are being provided as well with the understanding that all data and information will be made publicly available.

- the Systems Biology Group of the European Commission's Institute of Health and Consumer Protection (headed by Dr. Maurice Whelen) at the Joint Research Centre in Ispra, Italy, organized a two-day meeting in September 2010 to discuss potential collaborations with Tox21 and to sign an agreement with the U.S. EPA to screen the ToxCast™ Phase I 320 compounds in their HTS/high content screening facility.

- earlier this year, representatives of Health Canada attended a quarterly Tox21 general meeting in Research Triangle Park, NC, to learn more about our activities and to discuss future potential interactions.

- members of the U.S. Geological Survey met with NIEHS/NTP staff to discuss how the *in vitro* HTS methods we use to screen compounds for biological activity might be applied to water samples.

- recently, representatives of different toxicology laboratories with a public health function in the U.S. Department of Defense became members of the different Tox21 Working Groups, as a mechanism for being kept informed of our activities.

We fully appreciate that Tox21 faces some very difficult issues, such as (but not limited to):

- the lack of availability of HTS assays for measuring the free concentration of a compound *in vitro.*

- the lack of methods for the incorporation of xenobiotic metabolism into homogeneous HTS assays.

- the lack of HTS methods for evaluating interactions between cells and between tissues in response to single compounds and mixtures.

- the need to distinguish between statistical and biological significance.

- the difficulty in extrapolating from *in vitro* concentration to *in vivo* dose or blood levels.

- the difficulty in assessing the effects of chronic exposure conditions *in vitro*.

- most critically, how to identify when a perturbation to a gene/pathway would lead to an adverse effect in animals or humans.

The ultimate goal of our activities is to be able to integrate data from diverse technologies and endpoints into what is effectively a systems biology approach to toxicology. This can only be accomplished when comprehensive knowledge is obtained across a large portion of chemical and biological/toxicological space. The efforts described thus far reflect the initial stage of an exceedingly complicated program, one that will likely take decades to achieve its goals. However, even at this stage

of the process, the information obtained is impacting the international scientific community and the future of toxicology.

I am again pleased to acknowledge Dr. Christopher Portier for his vision in establishing and supporting the HTS program, the management of the NIEHS/NTP for their very active support of the HTS program, and the many individuals within the BSB, NTP, NIEHS, and the Tox21 partners for their commitment to a shared vision. The recent expansion of the BSB with regard to staff and, especially, the integration of the HSB into the BSB has greatly enhanced our ability to achieve our goals.